AGGLOMERATIVE FORCES AND CLUSTER SHAPES

William R. Kerr and Scott Duke Kominers*

Abstract—We model spatial clusters of similar firms. Our model highlights how agglomerative forces lead to localized, individual connections among firms, while interaction costs generate a defined distance over which attraction forces operate. Overlapping firm interactions yield agglomeration clusters that are much larger than the underlying agglomerative forces themselves. Empirically, we demonstrate that our model's assumptions are present in the structure of technology and labor flows within Silicon Valley. Our model further identifies how the lengths over which agglomerative forces operate influence the shapes and sizes of industrial clusters; we confirm these predictions using variations across patent technology clusters.

I. Introduction

AGGLOMERATION—industrial clustering—is a key feature of economic geography. A vast body of research now documents the prevalence of agglomeration in many industries and countries, and a number of studies have established agglomeration's particular importance for firm and worker productivity.[1] Moving from these measurements, researchers have recently sought to identify the economic rationales for firm collocation and thereby the sources of the associated productivity gains. While the list of potential suspects dates back to Marshall (1920)—most notably, labor market pooling, customer-supplier interactions, and knowledge flows—we are just beginning to separate the relative importance of these forces.

Research on the spatial horizons over which different agglomerative forces act often takes one of two approaches. A first approach considers regional evidence. Examples include Rosenthal and Strange (2001, 2004), Duranton and Overman (2005, 2008), and Ellison, Glaeser, and Kerr (2010). This approach begins by measuring the degree to which each industry is agglomerated across a chosen spatial horizon (e.g., counties, cities, states). A second step then correlates differences in observed agglomeration to the traits of the industries. For example, when using counties or cities as a spatial unit, we might observe that industries intensive in R&D efforts are more agglomerated than industries that do not depend on R&D. Similarly, we might observe that industries with strong customer-supplier linkages are agglomerated at the regional level. A common inference from these patterns, as one example, is that knowledge flows act over a shorter spatial distance than input-output interactions, as the knowledge-intensive industries are more heavily grouped together at the county level.

In parallel, a second strand of work considers local evidence on agglomerative interactions. Rather than discerning agglomerative forces from region–industry data, this line of research attempts to measure productivity gains directly at the establishment level. Prominent examples include Rosenthal and Strange (2003, 2008) and Arzaghi and Henderson (2008). A common approach is to estimate a plant-level production function that includes as explanatory variables the count of plants in the same industry observed within 5 miles of the focal plant, within 10 miles, and so on (or within the same county, city, and state). These studies often conclude that spillover effects decay sharply with distance, with the forces being orders of magnitude stronger over the first few city blocks than they are when firms are 2 to 5 miles apart. These productivity studies are just the tip of the iceberg, however, with many related research strands measuring directly the distances over which humans or firms interact (e.g., patent citations, commuting). As agglomerative forces depend on these interactions, these studies also describe the local interactions that give rise to the clusters that we observe.

There is a substantial gap between these two approaches. Despite their individual progress, we have very little understanding of how the local interactions aggregate up into regional shapes and sizes of industries we observe in the data. The easiest way to observe this gap is to consider the spatial distances discussed by the two approaches. The regional literature often concludes that technology spillovers have a shorter spatial horizon than labor market pooling by comparing county- and city-level data. But counties have populations over 75,000 people on average, and the spatial size of counties is much greater than what local interactions suggest is the relevant range. If studies find that knowledge flows decrease sharply within a single building (e.g., Olson & Olson 2003), why would we believe that we can infer useful comparisons of knowledge spillovers and labor pooling from

Received for publication May 28, 2013. Revision accepted for publication January 30, 2014. Editor: Philippe Aghion.

* Kerr: Harvard University and NBER; Kominers: Harvard University.

An earlier version of this paper was titled "Tipping Points and Agglomeration Bubbles." Alexis Brownell and Kristen Garner provided excellent research assistance. We thank David Albouy, Jerry Carlino, Jim Davis, Scott Dempwolf, Gilles Duranton, Ed Glaeser, Vernon Henderson, Bob Hunt, Guido Imbens, Yannis Ioannides, Sonia Jaffe, Ramana Nanda, Stuart Rosenthal, Steve Ross, Scott Stern, William Strange, and seminar participants at Aalto University School of Economics, American Economic Association, Bank of Finland, Boston University, Boston Urban and Real Estate Seminar, European Regional Science Association, Georgia Institute of Technology, Harvard University, NBER Productivity, NBER Urban Studies, Philadelphia Federal Reserve Board, Rimini Conference in Economics and Finance, University of California San Diego, University of Pennsylvania Wharton School, and University of Virginia for their insights. W.K. thanks the Bank of Finland for hosting him during a portion of this project. S.K. thanks the Becker Friedman Institute at the University of Chicago, the Harvard Center for Research on Computation and Society, the Harvard Program for Evolutionary Dynamics, and Microsoft Research New England for hosting him during parts of this project. The research in this paper was conducted while W.K. was a Special Sworn Status researcher of the U.S. Census Bureau at the Boston Census Research Data Center (BRDC). Support for this research from Harvard Business School (W.K. and S.D.K.); the Kauffman Foundation (W.K.); NSF grants ITR-0427889 (BRDC) CCF-1216095 (S.D.K.), and SES-1459912 (S.D.K.); an NSF Graduate Research Fellowship (S.D.K.); the Harvard Milton Fund (S.D.K.); and the Wu Fund for Big Data Analysis (S.D.K.) is gratefully acknowledged. Research results and conclusions expressed are our own and do not necessarily reflect the views of the Census Bureau or NSF. This paper has been screened to ensure that no confidential data are revealed.

A supplemental appendix is available online at http://www.mitpressjournals.org/doi/suppl/10.1162/REST_a_00471.

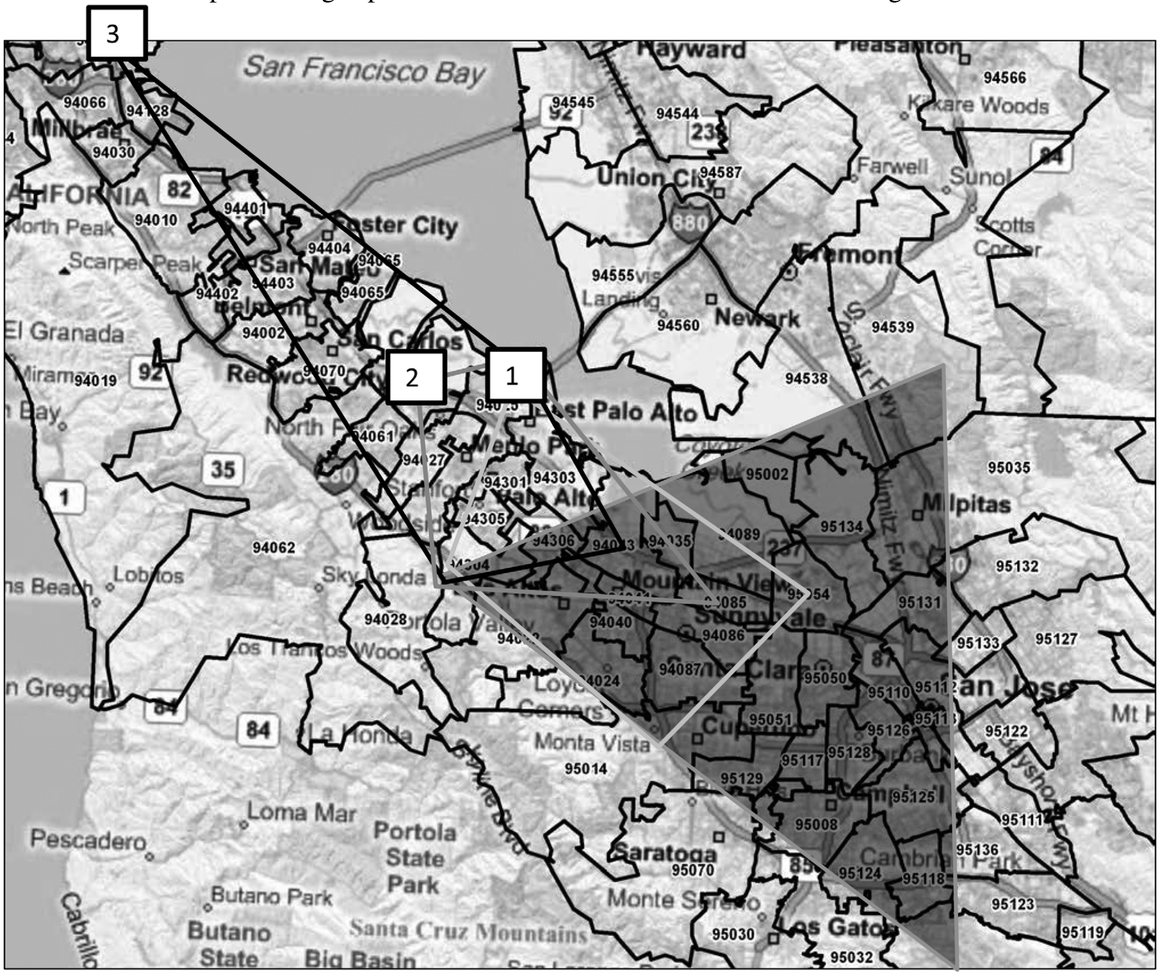
[1] Duranton and Puga (2004) and Rosenthal and Strange (2004) provide theoretical and empirical reviews, respectively.

The Review of Economics and Statistics, October 2015, 97(4): 877–899

doi:10.1162/REST_a_00471

FIGURE 1.—TECHNOLOGY SOURCING FROM SILICON VALLEY

Top Patenting Zip Codes outside of the Core and Their Sourcing Zones



Technology flows for the San Francisco area. The core of Silicon Valley is depicted with the shaded triangle. The Silicon Valley core contains 76% of the patenting for the San Francisco region. This map describes the technology sourcing for three of the four largest postal codes for patenting not included in the core itself. Technology sourcing zones are determined through patent citations.

The boxes indicate the focal postal codes, and the shape of each technology sourcing zone is determined by the three codes that firms in the focal postal code cite most in their work. Zone 1 for Menlo Park extends deepest into the core. Zone 2 for Redwood City shifts up and encompasses Palo Alto but less of the core. Zone 3 for South San Francisco further shifts out and brushes the core.

These technology zones are characterized by small, overlapping regions. None of the technology sourcing zones transverse the whole core, and only the technology zone of the closest postal code (Menlo Park) reaches far enough into the core to include the area of the core where the greatest number of patents occur. Transportation routes and geographic features influence the shapes and lengths of these sourcing zones.

The empirical appendix contains additional maps that show these small, overlapping regions are also evident in the core itself and in other areas outside the core.

regional data when the spatial scales of our data swamp the microinteractions by orders of magnitude?

This project examines these issues theoretically and empirically. The core of our work is a location choice model that connects limited, localized agglomerative forces with the formation of spatial clusters of similar firms. Agglomerative forces in our model are localized because firms face interaction costs. Spillover benefits exceed these interaction costs at short distances, and thus firms choose to interact. Beyond some distance, however, interaction becomes unprofitable and firms no longer engage with each other. For example, while a firm could learn useful technologies from another firm 20 miles away, the costs of doing so may be too great to justify the effort. Clusters are then the product of many small, overlapping regions of interaction. By building clusters up from microinteractions, we obtain additional insights into the structure of clusters and the regional data we observe.

Silicon Valley is the world's most famous cluster, and many observers credit its success to technology spillovers. Figure 1 illustrates the foundations of our theoretical framework using technology flows in Silicon Valley. Downtown San Francisco and Oakland are to the north and off the map. The triangle in the bottom right corner of the map is the core of Silicon Valley. This core contains 76% of industrial patents filed from the San Francisco Bay area and 18 of the top 25 postal codes in terms of patenting.

To introduce our model, we describe the primary technology sourcing zones for three of the four largest postal codes for patenting in the San Francisco area that are outside the core. Each focal postal code is marked with a box, and the other points of the shape are the three postal codes that firms in the focal postal code cite most in their work. Zone 1 for Menlo Park extends deepest into the core. Zone 2 for Redwood City shifts up and encompasses Menlo Park and Palo

Alto but less of the core. Zone 3 for South San Francisco further shifts out and brushes the core.

These technology zones are characterized by small, overlapping regions. None of the technology sourcing zones transverse the whole core, much less the whole cluster, and only the closest postal code (Menlo Park) even reaches far enough into the core to include the area of Silicon Valley where the greatest number of patents occur. While technology sourcing for individual firms is localized, the resulting cluster extends over a larger expanse of land.[2]

Our model replicates these features and makes explicit that empirical observation of cluster size in the data does not indicate the length of the microinteractions that produce the cluster. We show, however, that cluster shape and size does depend systematically on whether the localized interactions for firms in an industry are longer or shorter in length. We demonstrate that a longer effective spillover region, due to either weaker decay in benefits or lower interaction costs, yields a macrostructure with fewer, larger, and less dense clusters. These regularities allow researchers to use cluster dimensions to rank-order spillover lengths even though microinteractions are not observed. This connection helps bring together the diverse literature strands described earlier.

After deriving our theoretical predictions, we empirically illustrate the model using U.S. patent data to describe differences across technology clusters. Patent citations allow us to measure effective spillover regions by technology. Differences in these spillover regions relate to cluster shapes and sizes as predicted by the model. Technologies with very short distances over which firms interact exhibit clusters that are smaller and denser than technologies that allow for longer distances. This empirical work primarily employs agglomeration metrics that are continuous, as in the metric of Duranton and Overman (2005), and we use traits of industries in the United Kingdom to confirm the causal direction of these relationships (e.g., Ellison et al., 2010).

Our work makes several contributions to the literature on industrial agglomeration. Most important, we provide a theoretical connection between observable cluster shapes and the underlying agglomerative forces that cause them. Early island models of agglomeration, in which agglomerative forces act only within sites, implicitly feature maximal radius of interaction 0 (Krugman, 1991; Fujita & Thisse, 1996; Ellison & Glaeser, 1997). More recently, maximal radii also have been observed in more continuous models (Arzaghi & Henderson, 2008; Duranton & Overman, 2005). However, to our knowledge, our framework is the first to identify how variations in the maximal radius govern the shapes and sizes of clusters. At the core of this contribution is the simple mechanism of interaction costs among firms. The resulting framework provides a theoretical foundation for inferring properties of agglomerative forces through observed spatial concentrations of industries. We identify settings in which such inference is appropriate, as well as key properties of agglomeration in such settings.[3]

Our central empirical contribution is a framework, motivated by our theoretical model, for meaningful analysis of agglomerative forces with continuous distance horizons. Previous work considers how agglomerative forces affect spatial concentration over different distance horizons, for example, up to 75 or 250 miles (Rosenthal & Strange, 2001; Ellison et al., 2010). Our framework is an important step toward jointly considering agglomeration at different distances (25, 75, and 250 miles) simultaneously. We hope that future research can similarly analyze other factors that govern clusters' shapes and sizes.

In addition to the related work already mentioned, our empirical work with patents relates to two other recent studies that also consider continuous density measurements. Carlino et al. (2012) develop a multiscale core-cluster approach to measure the agglomeration of R&D laboratories across continuous space. In many respects, their metric's nesting approach parallels our theoretical focus on overlapping radii of interaction that build to a larger cluster. Likewise, some of their empirical results (e.g., clustering at local scales and at about 40 miles of distance) are also evident under our measures. Similarly, Murata et al. (2014) use continuous density estimations with patent citations to address the question of how localized are knowledge flows. Their careful metric design allows them to bridge the well-known debate between Jaffe, Trajtenberg, and Henderson (1993) and Jaffe, Trajtenberg, and Fogarty (2005) and Thompson and Fox-Kean (2005) and parse the underlying assumptions embedded in each study. Our work differs from these studies in several ways, but the most important difference is the theoretical focus and hypothesis testing about how different forms of interaction produce observable changes in cluster shapes and sizes.[4]

[2] The empirical appendix of our NBER working paper contains additional maps that show these small, overlapping regions are also evident in the core itself and in other parts of the San Francisco region. These properties are also evident in labor commuting patterns in the region. Arzaghi and Henderson (2008) and Carlino, Chatterjee, and Hunt (2012) provide related visual displays.

[3] An additional contribution of our work, discussed in greater detail later, is to provide a microfoundation for using continuous spatial density measurements that center on bilateral distances between firms. This class of metrics includes the popular Duranton and Overman (2005) metric.

Studies of agglomeration metrics include Ellison and Glaeser (1997), Maurel and Sédillot (1999), Marcon and Puech (2003), Mori, Nishikimi, and Smith (2005), Ellison et al. (2010), Billings and Johnson (2012), Carlino et al. (2012), and Barlet, Briant, and Crusson (2013). Recent related work on cities includes Helsley and Strange (2014) and Rozenfeld et al. (2011).

[4] Other related studies not previously mentioned include Audretsch and Feldman (1996), Ellison and Glaeser (1999), Head and Mayer (2004), Hanson (2005), Fallick, Fleischman, and Rebitzer (2006), Alcacer and Chung (2007), Aarland et al. (2007), Delgado, Porter, and Stern (2014), Pe'er and Vertinsky (2009), Holmes and Stevens (2002), Glaeser and Kerr (2009), Menon (2009), Alfaro and Chen (2010), Dauth (2010), Greenstone, Hornbeck, and Moretti (2010), Holmes and Lee (2012), Bleakley and Lin (2012), Dempwolf (2012), Marx and Singh (2012), Glaeser, Kerr, and Ponzetto (2013) and Helsley and Strange (1990). Our work also connects to studies of the shapes of cities (e.g., Lucas & Rossi-Hansberg, 2002; Baum-Snow, 2007, 2010; Glaeser, 2008; Saiz 2010) and of agglomeration and productivity differences across cities and regions (e.g., Ciccone & Hall, 1996; Partridge et al., 2009; Behrens, Duranton, & Robert-Nicoud,

Section II presents our theoretical model. Section III describes our empirical strategy and data and provides initial evidence for our model's building blocks using first- and later-generations of patent citations. Section IV then undertakes specific measurements of technology-level spillover radii and tests our model's predictions one at a time. Section V introduces our continuous density measurements and tests the model predictions. The last section concludes.

II. Theoretical Framework

We now introduce a model of firm location choice that generates large agglomeration clusters from smaller, overlapping spillover zones. To maintain consistency with previous work, we use the notation of Duranton and Overman (2005) whenever possible. We keep this initial exposition as simple as possible and conclude this section with a discussion of richer frameworks and extensions.

A. Basic Framework

There are N firms indexed by i. These firms i sequentially select their locations, denoted $j(i)$, from a fixed set $\mathbf{Z} \subset \mathbb{R}^2$ of potential sites, each of which can hold at most one firm.[5] Sites are drawn at random according to a uniform distribution in advance of any firm's location decision. There are many more possible sites than firms: $|\mathbf{Z}| \gg N$. To focus on agglomeration economies, we assume that firms compete in broad product markets. Location choice thus affects the productivity of a firm, but not its competitive environment.

The specific benefits of location j to a firm are driven by intraindustry Marshallian forces representing productivity spillovers that firms generate by being in proximity to each other. Three common examples are customer-supplier interactions (e.g., reducing transportation costs for intermediate goods), labor pooling, and knowledge exchanges.

We denote by d_{j_1,j_2} the spatial distance between $j_1 \in \mathbb{R}^2$ and $j_2 \in \mathbb{R}^2$. We assume that the deterministic benefit of site $j \in \mathbf{Z}$ to a firm i is given by

$$g_j(i) \equiv \sum_{i' \neq i} G(d_{j,j(i')}),$$

for some continuous, decreasing function G. The value g_j represents the degree to which spillovers from other sites make site j specifically attractive to firms. We assume the standard comparative static that G is decreasing, so that agglomerative forces decline over space. Additionally, for simplicity, we assume that agglomerative forces act across all distances: that is, $G(d) > 0$ for all $d \geq 0$.

We assume that a firm chooses randomly among sites $j_1, \ldots, j_\ell$ over which that firm would be indifferent if forced to choose purely on the basis of spatial attraction.[6] We also assume that firms are not forward looking, so that the nth firm to enter, i_n $(1 \leq n \leq N)$, chooses its location $j = j(i_n) \in \mathbf{Z}$ to maximize $g_j(i_n)$ conditional on the location choices of the first $n - 1$ firms.

B. Maximal Radii of Interaction

So far, our model has more or less followed a standard structure: proximity to resources and other firms generates benefits, and these benefits decay continuously over distance. However, we now depart from this standard approach via a simple and natural additional assumption:

Assumption. *A firm must pay cost $c > 0$ to interact with another firm.*

The fixed costs c relate to the costs of transporting goods, people, or ideas across firms. Opportunity costs and search costs are the simplest examples, and these costs can be specific to industries and spillover types. For example, accessing and understanding codified technologies likely requires a lower fixed cost of establishing interactions than that required for tacit technologies.[7]

Firms invest in establishing contacts when the benefits of doing so equal or exceed the associated costs of interaction. Specifically, firm i invests in contact with a firm i' only if $G(d_{j(i),j(i')}) \geq c$. This defines a strict distance over which firm i finds interactions profitable:

$$d_{j(i),j(i')} \leq \rho \equiv \max\{d : G(d) \geq c\}.$$

Therefore, we immediately observe the following result:

Proposition 1. *Firms at sites farther than distance ρ from a firm i cannot profitably interact with i. That is, firm i derives no direct benefits from the presence of firms at locations j with $d_{j,j(i)} > \rho$.*

Proof. Immediate from text.

The key consequence of proposition 1 is that agglomerative forces in practice act only over finite distances. We call ρ the maximal radius of interaction (or just the maximal radius). The maximal radius is (weakly) decreasing in the cost c and increasing in the levels of the decay function G. In other words, lower costs or weaker attenuation of benefits lead to larger maximal radii.

Our assumption that interaction costs are fixed is only to simplify the discussion that follows. One might naturally assume that interaction costs rise with distance; such

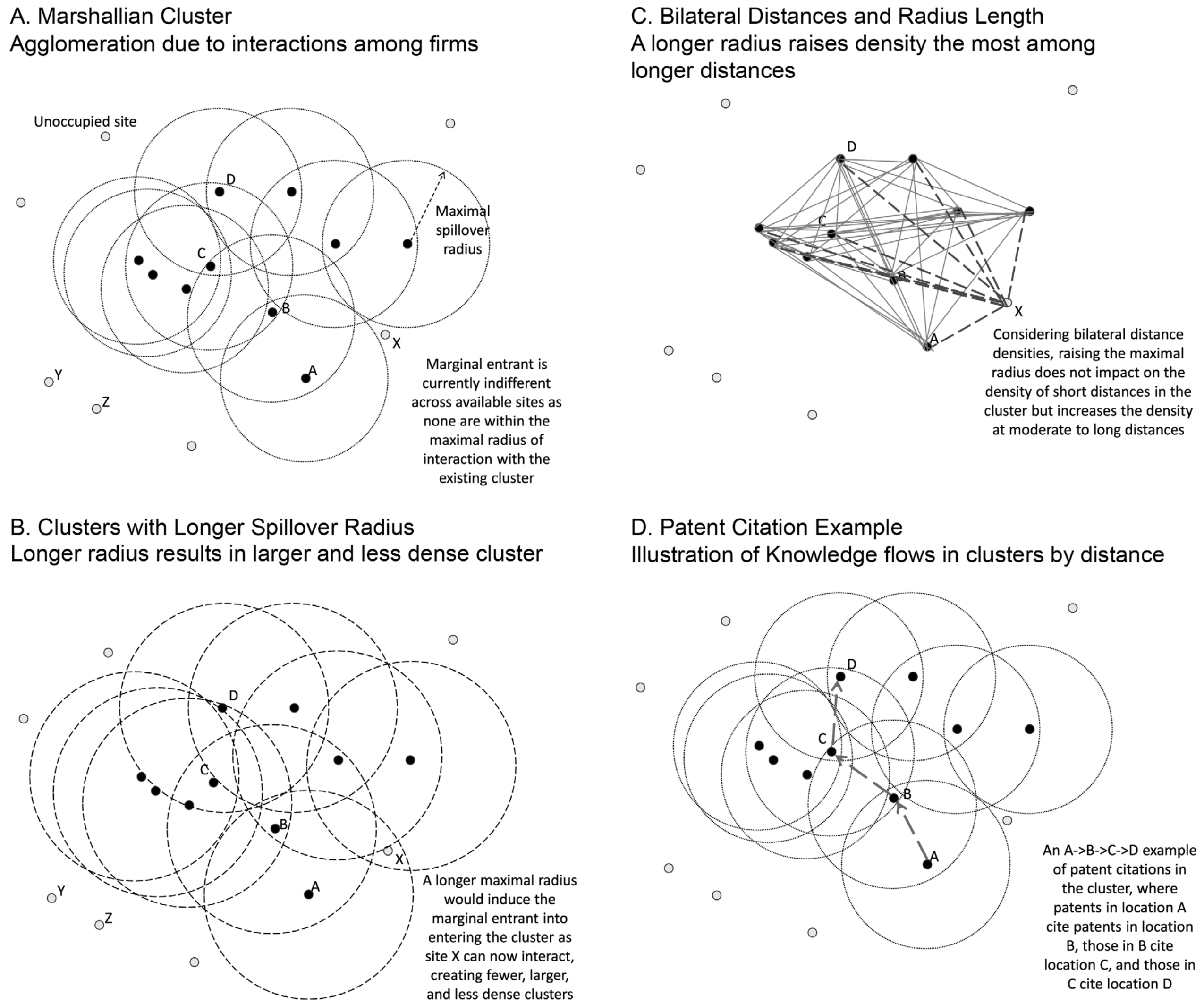
2014; Sarvimäki, 2010; Fu & Ross, 2013). Jackson (2008) outlines a complementary literature on economic networks.

[5] In section IID, we discuss the possibility that multiple firms may occupy (and congest) the same site.

[6] The exact specification of the distribution of random site choice does not matter for our theoretical results and may be conditioned on the set of sites already occupied, but we do require that it is identical across firms.

[7] Arzaghi and Henderson (2008) use a similar foundation in their model of location choice for ad agencies in Manhattan.

FIGURE 2.—ILLUSTRATION OF THE THEORY MODEL



A. Image illustrates a Marshallian cluster. Entry is sequential, without foresight, and potential sites are fixed. Black dots are chosen sites, and circles represent maximal spillover radii. Spillover radii are limited due to fixed costs of interaction. Large area clustering is due to small, contained interaction effects that overlap each other. The next entrant is indifferent among available sites, including sites X, Y, and Z.

B. Illustrates a cluster for an agglomeration force with a longer maximal radius. The larger dashed circles show that a longer maximal radius would induce the marginal entrant into the cluster at site X over the other sites, resulting in (weakly) larger and less dense clusters. An additional prediction is that there should be fewer clusters for a technology given a fixed number of firms.

C. The light gray lines show bilateral connections for the cluster formed under the shorter maximal radius in panel A. The large, dashed lines show the additional bilateral connections formed when the cluster grows due to the longer maximal radius of interaction in panel B. The length of the bilateral connections from the induced entry at site X will be longer than the shortest existing bilateral connections.

D. The image provides an example of knowledge flows within a cluster discussed in the text. Knowledge is flowing from site D to site C to site B to site A, and we thus observe patent citations in the reverse order. This example is contrived so that site A will interact only with site D under the given maximal radius via interconnections provided by other sites.

an assumption would also generate the maximal radius described in proposition 1. The ultimate technical condition required is that interaction costs exceed interaction benefits at some distance with a single crossing.

C. *A Cluster-Based Theory of Agglomeration*

We next examine how clusters form in our model and illustrate clusters' properties. Figures 2A to 2D provide a graphical presentation of the theory to build intuition. In these graphs, lightly shaded circles are potential firm locations, and filled-in circles represent sites populated by firms. Throughout this paper, we use these graphs to explain the model's structure and depict the behavior of marginal entrants.

Basic definitions and structure. We define an *agglomeration cluster* to be a group of firms located in sites interconnected by bilateral interactions. Each firm does not necessarily interact with every other firm in its cluster, but all firms in a cluster are interconnected. Our measure of agglomeration counts the number of these clusters that are expected to arise; we say that firms exhibit agglomeration if they typically occupy few distinct clusters. (The use of an expectation is necessary because firms choose randomly when indifferent among sites.)

More formally, for $j \in \mathbb{R}^2$, we denote by $B_d(j) \equiv \{j' \in \mathbb{R}^2 : d_{j,j'} \leq d\}$ the closed ball of radius d about j. For $j \in \mathbf{Z}$, we set

$$\mathcal{B}^0_\rho(j) \equiv B_\rho(j) \cap \mathbf{Z}.$$

This formula has a simple interpretation: $\mathcal{B}^0_\rho(j)$ is the set of potential firm locations that can profitably interact with j under maximal radius ρ.

In figure 2A, we draw for each populated site a representative maximal radius within which the benefits of interaction exceed the costs for firms. For this example, $\mathcal{B}^0_\rho$ for site B includes sites A and C. Sites A and C are the only locations within the maximal radius for Marshallian spillovers ρ.

We next expand our focus to consider sites that are outside the profitable spillover range of site j but can be connected to j with a single interconnection. We define $\mathcal{B}^1_\rho(j)$ to be the set of sites that can profitably interact with the sites in $\mathcal{B}^0_\rho(j)$ through one additional step. In figure 2A, $\mathcal{B}^1_\rho(\mathrm{B})$ further has the four additional sites within distance ρ from site C that are outside the spillover range of site B. We continue to iterate this process, successively adding sites that are more spatially distant to site j but still connected to site j by increasing numbers of interconnections ($\mathcal{B}^\iota_\rho(j)$ for $\iota = 1, 2, \ldots$). Formally, for any $j \in \mathbf{Z}$,

$$\mathcal{B}^\iota_\rho(j) \equiv \bigcup_{j' \in \mathcal{B}^{\iota-1}_\rho(j)} \mathcal{B}^0_\rho(j').$$

Iterating this construction of clusters to its conclusion, $\mathcal{B}_\rho(j)$ is the ρ-cluster containing $j \in \mathbf{Z}$, defined by

$$\mathcal{B}_\rho(j) \equiv \bigcup_{\iota=0}^{\infty} \mathcal{B}^\iota_\rho(j).$$

The ρ-cluster containing site j is the largest cluster of sites that contains j and is connected by a chain of "hops" between sites $j' \in \mathbf{Z}$ that can profitably interact. The complete set of filled locations in figure 2A constitutes the ρ-cluster for site B in our example. The use of an infinite index in the union defining $\mathcal{B}_\rho(j)$ is ultimately unnecessary, as the finitude of the set of sites implies that $\mathcal{B}^\iota_\rho(j) = \mathcal{B}^{\iota+1}_\rho(j) = \cdots$ for some finite ι.

When the maximal radius ρ is small, clusters are generally small. For two precise examples, define the lower and upper bounds on distances between sites as $\underline{d} \equiv \min_{j_1 \neq j_2 \in \mathbf{Z}} d_{j_1,j_2}$ and $\bar{d} \equiv \max_{j_1 \neq j_2 \in \mathbf{Z}} d_{j_1,j_2}$. When $\rho < \underline{d}$, $\mathcal{B}_\rho(j) = \{j\}$ for each $j \in \mathbf{Z}$. In other words, a maximal radius that is shorter than the shortest distance between two sites results in each cluster containing only a single firm. By contrast, when $\rho > \bar{d}$, we have $\mathcal{B}_\rho(j) = \mathbf{Z}$ for all $j \in \mathbf{Z}$—a maximal radius longer than the maximal distance between sites results in a single cluster for an industry.

If the maximal radius is ρ and the first firm locates at site j, and the cluster around j contains available locations (i.e., $\mathcal{B}_\rho(j) \neq \{j\}$), then there is some site $j' \in \mathcal{B}_\rho(j)$ that delivers positive deterministic utility flows to the next entrant. It follows that if Marshallian forces are sufficiently strong, firms select sites in the cluster $\mathcal{B}_\rho(j)$ until $\mathcal{B}_\rho(j)$ is filled. Iterating this analysis shows that when Marshallian forces are strong, firms fill clusters sequentially.

The sequential filling of clusters explains how large-area clustering may arise in an industry even if agglomerative forces act only over short distances. Cluster sizes associated with a given maximal radius can be much larger than the underlying radius itself. Clusters may span large regions even if each firm derives benefits only from its immediate neighbors.

A consequence of the maximal radius, however, is that clusters can reach their capacity, at which point the next entrant for the industry will locate elsewhere. In figure 2A, the closest remaining site to the existing cluster is site X, but this location is beyond the spillover ranges of any of the populated sites in the cluster. Because the marginal entrant cannot profitably interact with the cluster, it is indifferent among sites X, Y, Z, and any other unoccupied site. It will choose its location at random or based on idiosyncratic preferences.

These observations suggest a natural notion of agglomeration. We say that firms are (weakly) more agglomerated with respect to maximal radius ρ_1 than they are with respect to radius ρ_2 if, holding N fixed, fewer clusters of firms form when the maximal radius is ρ_1 than when it is ρ_2. Formally:

Definition 1. *The level of agglomeration for maximal radius ρ is $|\mathbf{Z}| - \Xi_\rho$, where Ξ_ρ is the expected number of distinct ρ-clusters $\mathcal{B}_\rho(j)$ about sites $j \in \mathbf{Z}$ occupied by firms.*

Note that under this definition, agglomeration increases as the expected number of clusters decreases. When industry size is held constant, increased agglomeration also corresponds to increased cluster size. The additive term $|\mathbf{Z}|$ is a normalization that guarantees that the level of agglomeration is always a positive number. We could equally well define the level of agglomeration for maximal radius ρ to just be $-\Xi_\rho$.

Our discussion of the marginal entry decision also highlights the core difference between our structure and prior work. Without considering interaction costs, strictly positive spillover benefits exist at all distances due to the decay function G. Industries may differ in how fast or slow Marshallian benefits decay, but these differences in Marshallian forces do not affect the number of clusters. Regardless of whether the potential spillover benefit is large or minuscule, marginal entrants always select sites closest to the developing cluster regardless of distance (thus, in figure 2A, site X is chosen next). As a consequence, each industry always forms a single cluster, and entrants generically select sites in a fixed order. This is equivalent to the case in which $\rho \to \infty$ in our model. Once a first entrant picks a location, the set of sites filled by the remaining $N - 1$ firms is exactly determined.

Thus, the simplest framework does not provide a foundation for relating differences in spatial concentration for industries to their underlying agglomerative forces. Yet our intuitive addition of interaction costs provides additional traction by establishing a spatial range over which interactions are relevant. This localization in turn provides meaningful differences in cluster formation. We now turn to these comparative statics.

Agglomeration due to Marshallian advantages. Figure 2B illustrates the consequences of a longer maximal radius for cluster formation. Under the larger maximal radius, the marginal entrant is no longer indifferent over sites but would instead choose site X. Thus, a longer maximal radius is (weakly) associated with greater industry agglomeration as fewer clusters form in expectation.

More formally, recall that ρ is the maximal radius for intraindustry spillovers, $\rho \equiv \max\{d : G(d) \geq c\}$. Since $\mathbf{Z}$ is finite, small changes in ρ do not affect location choices. Larger increases in ρ, due to either weaker attenuation in spillover benefits or lower interaction costs, can lead firms to organize into fewer clusters. In fact, we may sign this change: firms become (weakly) more agglomerated when ρ increases.

Proposition 2. *A longer maximal radius of intraindustry spillovers* ρ *leads to a (weak) increase in agglomeration.*

Proof. See the online appendix.

The idea behind the proof of proposition 2 is intuitive. A firm i that is indifferent across sites chooses its location $j(i) \in \mathbf{Z}$ randomly. But until the sites in cluster $\mathcal{B}_\rho(j(i))$ are filled, they are more attractive to firms than are unfilled sites outside $\mathcal{B}_\rho(j(i))$. If ρ grows to $\hat{\rho}$, a radius large enough to cause some cluster $\mathcal{B}_\rho(j)$ to merge with another cluster (i.e., such that $\mathcal{B}_{\hat{\rho}}(j) = \mathcal{B}_\rho(j) \cup \mathcal{B}_\rho(j')$ for some site $j' \notin \mathcal{B}_\rho(j)$), then the expected number of clusters occupied by firms shrinks. Indeed, whenever a firm locates in either $\mathcal{B}_\rho(j)$ or $\mathcal{B}_\rho(j')$, subsequent firms fill all of $\mathcal{B}_{\hat{\rho}}(j)$ before locating in or starting another cluster.

Three empirical implications of this analysis are evident in figure 2B. First, industries with a longer maximal radius have larger clusters in the sense of having more firms and covering a greater spatial area. Intuitively, a longer spillover radius makes sites at the edges of clusters attractive, even if they would not be attractive given a shorter radius. This induces marginal entrants into choosing these sites rather than starting new clusters. A longer radius can be due to weaker decay of spillover benefits or lower interaction costs.

The second and third predictions are closely related. A longer spillover radius yields fewer clusters for a given industry size. As clusters grow in size, fewer clusters are needed to house the N firms in the industry. Finally, clusters are less dense. The longer radius activates sites at the edges of a cluster that are too spatially distant to profitably interact with previous entrants if the radius is shorter. Thus, growth in cluster size is simultaneous with reduction in cluster density.

Our result that clusters due to a longer maximal radius are less dense is the same as saying that average bilateral distances among firms within the clusters increase. The model's structure, however, contains a much more powerful implication regarding spillover lengths and the complete distribution of bilateral distances within clusters. We draw out this implication below.

Ordering and characterizing agglomerative forces. The theory suggests that longer maximal spillover radii are associated with fewer, larger, and less dense clusters. It is feasible to use these observed traits in different industries to rank-order the radii associated with different spillovers. In the empirical analyses that follow, for example, we provide suggestive evidence for the model by plotting an estimate of the maximal radii for different technologies against a measure of technology cluster density. This section introduces terminology and conditions required to jointly test these predictions using the continuous density estimation techniques employed in section V.

It is impossible to measure directly the G functions that determine the value of firm clustering. However, observed spatial location patterns allow us to partially model the behavior of the unobserved functions G in a continuous manner.

Proposition 3. *Holding* ρ *fixed, and assuming that* G *is differentiable, an increase in* $|G'|$ *leads to a (weak) increase in the number of firms clustered at small distances.*

Proof. Immediate from text.

The decay of agglomerative forces across space correlates with observed distances between clustered firms. Thus, we may understand the speed at which the benefits of localization decay by measuring the degree of localization at different distances. For an extreme example, if localization of firms is constant across space, then we must have $|G'| = 0$. If localization gradients are very sharp at short distances, then proposition 3 implies that the underlying G function sharply attenuates. Note that intercept value $G(0)$ is not held fixed in proposition 3. An implication of our framework is that, holding ρ fixed, $G(0)$ affects the gradient $|G'|$, but does not affect the overall level of agglomeration.

Proposition 3 allows us to use Duranton and Overman's (2005) density estimations in section V to characterize distributions continuously. Adding this more continuous structure to our model, we can compare the full distributions of industries to assess how longer maximal radii affect the shapes of clusters. The predictions that clusters become larger and less dense become jointly visible. Moreover, we can observe this effect's influence using regular step sizes in distance.

Let $\mathbf{S}$ denote the set of sites occupied by firms in equilibrium, with many industries present in the economy. The null hypothesis is that neither localization nor dispersion occurs when the maximal radius is ρ, that is, $g_j = 0$—firms locate randomly—when the maximal radius is ρ. We empirically proxy the set of potential sites $\mathbf{Z}$ with the observed set of actual sites $\mathbf{S}$ for all businesses. With this assumption, density measures can quantify localization by comparing observed localization levels to counterfactuals representing the underlying distribution of economic activity typical for a bilateral distance. The null hypothesis is rejected if the localized density of firms is a substantial departure from counterfactuals having (the same number of) firms occupying sites randomly sampled from **S**.[8]

Bilateral distance gradients in agglomeration clusters. There is an additional benefit to connecting our model to these continuous structures. We earlier noted that our empirical implication of smaller, denser clusters for a shorter maximal radius is equivalent to saying that the mean bilateral density for clusters declines. The model, however, has a stronger implication for how spillover length influences the distribution of bilateral distances within clusters.

Proposition 4. *There is some $\bar{\rho} > \underline{d}$ so that whenever ρ and ρ' are such that $\underline{d} < \rho < \rho' < \bar{\rho}$, then the mean intracluster firm distance is (weakly) smaller when the maximal radius is ρ than when it is ρ'.*

Proof. See the online appendix.

This result describes a key comparative static across spillover lengths. When comparing two industries, we earlier established that the industry with the shorter maximal radius should exhibit denser clusters such that very close bilateral distances are common. This proposition further identifies that this greater representation should be at its highest at the shortest bilateral distances possible (i.e., among locations very near to each other). This higher frequency should then (weakly) decline as one considers bilateral distances farther from the shortest possible connections.[9]

To provide intuition, first consider the impact of the marginal entrant on the bilateral distances in figure 2C. As site X becomes part of the cluster, the set of bilateral distances grows to incorporate the bilateral distance from site X to every other populated site in the cluster into the spatial description. Some of the added bilateral distances are shorter than those that already existed in the cluster, with the distance between sites X and B, for example, being less than the distance between sites A and D. Yet all of the additional bilateral distances are longer than the closest connections possible (e.g., those surrounding site C). Thus, as the cluster expands and becomes less dense, the relative impact on densities is most at the shortest possible connections and proceeds (weakly) outward for some distance.

An empirical example can also help. Assume that the premium for proximity is higher for investment bankers than it is for accountants. We predict that clusters of investment bankers should exhibit shorter mean bilateral distances among firms than clusters of accountants do. When comparing the spatial distributions of their clusters, proposition 4 further indicates that the greater density for investment banking should be at its highest at the spatial level of being in the same building or on the same city block. When looking at firms five blocks away from each other, the spatial density for investment bankers can still exceed that of accountants, but the difference should not be higher than it is when looking at firms next door to each other.

This requirement microfounds use of continuous density metrics like that of Duranton and Overman (2005) in assessing whether differences in agglomerative forces across industries yield meaningful deviations in agglomeration behavior. To summarize, we should empirically see that the greater density associated with a shorter maximal radius is at its maximum at the closest possible distance on the spatial scale and (weakly) declines thereafter for some distance. Eventually a distance is reached where the bilateral densities are the same even with the differences in maximal radius. Continuing with our earlier example, the offices of investment bankers and accountants may be equally represented when looking at firms that are ten city blocks apart.

After this point, a distance interval follows with relative underrepresentation for the cluster associated with the shorter maximal radius. Finally, once spatial distances are reached that represent distances between agglomeration clusters for Marshallian industries, the relative densities again converge. In our example, accounting firms should be more represented than investment bankers when looking at businesses fifteen to twenty blocks from each other. This higher representation of accountants should then decline as we consider progressively longer distances that start to exceed the sizes of cities.

By contrast, our model generally does not make predictions for bilateral distances across Marshallian industries beyond the spatial horizons of individual clusters. The behavior of longer horizons depends on the underlying distribution of cluster sites and it is thus ambiguous in our framework. The median bilateral distance for all firms within an industry, for example, can increase or decrease with a longer maximal radius depending on the spatial distances among the multiple, growing clusters and the newly activated sites surrounding them.

D. *Discussion*

We now discuss potential enrichments of the model. We first note that this model is a simplified version of the one contained in our NBER working paper. The version we present here assumes that all firms belong to the same industry. We also abstract away from the possibility of clustering due to fixed, location-specific natural advantages (e.g., coal mines, universities). The extended theoretical framework relaxes both of these simplifications and shows that they do not materially affect the predictions for Marshallian clusters that we develop and test here. Our NBER working paper also outlines some basic spatial dynamics for clusters.

For simplicity, our base model allows at most one firm per site. Our results are unchanged if we allow multiple firms to

[8] As discussed in the empirical appendix of our NBER working paper and in Barlet et al. (2013), this approach is slightly strained for the largest industries but is a reasonable baseline for most industries.

[9] The conditions of proposition 4 indicate that this effect may disappear when the maximal radius is very large. This is a natural consequence of approaching a limit where the maximal radius is so large as to no longer influence cluster formation.

locate at each site and assume that collocated firms "congest" each other. Specifically, we may extend our model by assuming that each site $j \in \mathbf{Z}$ has a maximum capacity $\kappa_j \geq 1$ and that the set $I(j)$ of firms located at j must always have $|I(j)| \leq \kappa_j$. Congestion is modeled by assuming that firms $i \in I(j) \neq \emptyset$ receive spillover benefits of $\hat{g}_j(i) \equiv (1/|I(j)|) \cdot g_j(i)$. That is, spillovers to location j are divided equally among firms collocated at j. With these notations, our base model corresponds to the case where $\kappa_j = 1$ for all sites $j \in \mathbf{Z}$. Even with congestion, the sequential location choice model is justified: a firm i entering site $j(i)$ has the potential to "crowd" its closest neighbors. But that firm i can never crowd out another firm $i' \in I(j(i))$. Indeed, if firm $i' \in I(j(i))$ were to exit $j(i') = j(i)$ and relocate after the entrance of firm i, then on relocation, i' would face the same location choice problem previously faced by firm i. The ex ante optimality of $j(i)$ for i would then show that $j(i)$ is the ex post optimal choice for i'.

Second, the microinteractions across sites that are built into the model are readily generalized. Our discussion and proofs focus on the simple case where spillover benefits do not transfer through the cluster. Interaction costs are incurred on a bilateral basis, and firms at the periphery of a cluster receive benefits only from their immediate neighbors. More generally, our predictions hold for any structure of benefit transmission through the cluster so long as ρ is constant, as the spillover radius at the cluster's edge is what determines the marginal entrant's decision. We might also assume that with some probability $p(d) < 1$, firms invest in contact with firms of distance d away, with p declining in d (i.e., $p'(d) < 0$). With our model's structure, we can handle this case by simply replacing the function $G(d)$ with $G(d) \times p(d)$. Alternatively, if there is always some (possibly small) fixed probability that a firm chooses its location randomly, as in the model of Ellison and Glaeser (1997), then our qualitative conclusions are maintained: an upward shift in p leads to fewer, larger, and less dense clusters.

Finally, the model does not include property prices. One way to introduce property prices to the model is to consider them as the consequence of wanting to be near a fixed feature (e.g., the city center). The version of the model in our NBER working paper shows that this extension does not materially affect our predictions for Marshallian agglomeration so long as feature attraction effects are not too strong.

III. Patent Technology Clusters

A. *Overview of Empirical Strategy*

We illustrate the model's predictions empirically in this section using variation across patent clusters. We proceed in three steps that closely follow the model's structure. We first use patent citation data to illustrate how knowledge flows within U.S. technology clusters resemble the model's maximal radius construct. Patent citations provide a rare window into the distances over which knowledge interactions and technology flows are occurring within clusters. In a generalization of the Silicon Valley case study in section I, we demonstrate how these knowledge flows are limited in distance even within a single cluster. We also show how bilateral interactions form overlapping regions of interaction that cover a larger spatial area than the individual interactions of firms do.

After establishing these properties generally, we use the patent data to calculate differences in the lengths of maximal radii across technology groups. Some technology areas (e.g., semiconductors) have very localized citation patterns where knowledge flows decay rapidly with distance. Knowledge flows in other technology areas operate effectively over longer distances. After measuring these differences across technologies, we turn to our basic model predictions that a longer radius of interaction generates larger and less dense clusters (propositions 1 and 2), showing that each of these basic predictions holds when considered independently. We do not investigate the number of clusters prediction, as it is substantially more sensitive to empirical choices than the properties of clusters are.

Our final exercises present a unified empirical framework for analyzing how technology cluster shapes and sizes differ across technologies in relation to their maximal radii of interaction. This framework brings to bear the joint nature of our three main predictions and the more subtle predictions of propositions 3 and 4 with respect to rates of relative decay. These tests require that we depict the whole distribution of distances within a cluster and analyze the differences in these shapes across technologies. We conduct these tests using a mixture of nonstructured plots and the continuous spatial density metrics developed by Duranton and Overman (2005). These depictions provide greater insights into how observable cluster shapes provide information about the underlying agglomeration force.[10]

B. *Patent Citations and Knowledge Flows*

We employ individual records of patents granted by the U.S. Patent and Trademark Office (USPTO) from January 1975 to May 2009. Each patent record provides information about the invention (e.g., technology classification, firm or institution) and the inventors submitting the application (e.g., name, address). Hall, Jaffe, and Trajtenberg (2001) provide extensive details about these data, and Griliches (1990) surveys the use of patents as economic indicators of technology advancement. The data are extensive, with over 8 million

[10] This section's investigation most closely relates to knowledge flows as a rationale for agglomeration and cluster formation. Section IV of our NBER working paper provides additional empirical evidence for the model's structure when comparing the distances over which knowledge flows occur to distances over which agglomeration is driven by labor pooling or natural advantages. These supplementary exercises have the advantage of covering many industries and sectors in the U.S. economy, but the broader approach means that we no longer identify the microinteractions among firms as we do in patent data. What we show is that the ordering of industries by these various agglomeration rationales produces patterns in line with our model.

inventors and 4 million granted patents during the data period.

A long literature exploits patent citations to measure knowledge diffusion or spillovers. A number of studies examine the importance of local proximity for scientific exchanges, generally finding that spatial proximity is an important determinant of knowledge flows.[11] Additional work links these local exchanges and economic clusters. Carlino et al. (2007) find that higher urban employment density is correlated with greater patenting per capita within cities. Rosenthal and Strange (2003) and Ellison et al. (2010) find that intellectual spillovers are strongest at the very local levels of proximity. These empirical patterns closely link to ethnographic accounts of economic activity within clusters (Saxenian, 1994).[12]

Patent citations thus offer us a unique opportunity to quantify differences in spillover radii and cluster shapes. It is important, however, to recall several boundaries of this approach. First, patent citations can reasonably proxy for technology exchanges, but there are many other forms of knowledge spillovers that may behave differently (Glaeser & Kahn, 2001; Arzaghi & Henderson, 2008). Second, several studies find that patent citations reflect Marshallian spillovers among firms other than pure knowledge exchange. Breschi and Lissoni (2009) closely link citations to inventor mobility across neighboring firms in their sample, and Porter (1990) emphasizes how technologies embodied in products and machinery can be transferred directly through customer-supplier exchanges. Our measurements below may encompass these effects to the extent that they operate.

C. *Patent Data Construction*

Inventors are required to cite the prior work on which their patents build. The total count of citations made by USPTO domestic and foreign patents granted after 1975 is about 41 million. We first restrict this sample to citations where the citing and cited patents are both applied for after 1975. This restriction is necessary for collecting inventor addresses. Our second restriction is that both patents have inventors resident in the United States at the time of the invention with identifiable cities or postal codes. About 15 million citations remain after these restrictions. Our primary data set further focuses on the 4.3 million citations that are made in a geographical radius of 250 miles or shorter from the citing patent.

To identify these distances, we extract postal codes from addresses given for inventors. This data set combines both postal codes listed directly on patents and representative postal codes taken from city addresses where postal codes are not listed. Where multiple inventors exist for a patent, we take the most frequent postal code; ties are further broken using the order of inventors listed on the patent. The spatial radius is defined using geographic centroids of postal codes and the Haversine flat earth formula. We assign a distance of less than 1 mile to cases where the citing and cited patents are in the same postal code.

Our analyses consider how distances between postal codes influence patent citation rates. Several issues with using inventor postal codes should be noted. A small concern is that our approach does not consider all of the postal codes associated with inventors for some patents, and this may lead to mismeasurement in our distance measure over short spatial scales (specifically, an upward bias on the minimum distance). As a check against this concern, we find very similar results when instead employing only patents with single inventors. More substantively, listed addresses can represent either home or work addresses. It would be nice to model both distances between work locations and distances between inventor home locations. Both of these distances can influence technology diffusion, and it is not clear which is more important. The patent data do not let us separate these two, however, and this measurement error biases us against finding shorter spillover effects.

To ensure that our results are not overly dependent on our approach, especially with respect to the maximum radii that we calculate by technology, we also calculate a parallel set of distances using a match of USPTO patents to firms in the Census Bureau (Kerr & Fu, 2008; Balasubramanian & Sivadasan, 2011; Akcigit & Kerr, 2010). The Census Bureau data records identify the postal codes of each firm's establishments in a city. We thus take the patents identified to be in Chicago for a particular firm, for example, and assign them the postal codes of the firm's records. Unreported analyses confirm the spillover radii that we identify with our primary data set.[13]

D. *Knowledge Flows within Clusters*

Our first analysis characterizes how knowledge flows within technology clusters. To do so, we examine patent citation patterns, specifically differences in spatial scope within clusters for first-generation citations compared to later

[11] See Jaffe, Trajtenberg, and Henderson (1993), Jaffe, Trajtenberg, and Fogarty (2000), Thompson and Fox-Kean (2005), Thompson (2006), and Lychagin et al. (2010). Murata et al. (2014) measure the continuous density of patent citations.

[12] Recent theoretical and empirical work further ties innovation breakthroughs to the clustering of activity around the discovery location, suggestive of very short spillover ranges (Zucker, Darby, & Brewer, 1998; Duranton, 2007; Kerr, 2010). These concepts are central to endogenous growth theory (e.g., Romer, 1986), and Desmet and Rossi-Hansberg (2014) present a recent model of spatial endogenous growth.

[13] The primary advantage of the work using the Census Bureau's data is to verify robustness with a second data source. There are two disadvantages. First, we would need to disclose any results that we wish to report using the Census Bureau's data. Basing our primary estimations on inventor address data outside the Census Bureau allows us much more flexibility for generating graphs of continuous density estimates. Second, the Silicon Valley case study in section I (where we manually identified postal codes for work locations) was attractive in that single firm locations typically house both corporate headquarters and innovation facilities. This collocation is much less prevalent in the New York City region, for example, where major firms frequently have offices in Manhattan and in surrounding areas. These multiple offices even within 250-mile circles limit the gain from using establishment-based identifiers versus simply using known inventor addresses.

generations of citations. This analysis is useful because it provides evidence of the interconnections among firms built into our model's structure. It also introduces the empirical framework that we use to calculate the maximal radius for each technology.

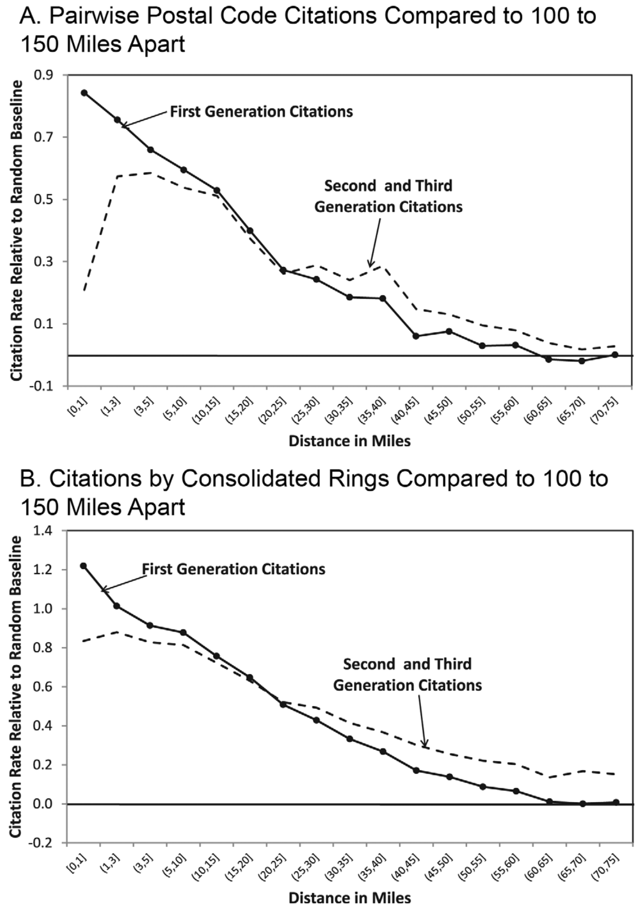
To introduce and clarify terminology, consider a sequence of patents where patent A cites patent B, patent B cites patent C, and patent C cites patent D. Using an arrow to indicate a citation, our sequence is A→B→C→D. Note that in this example, the citations are moving from patent A to patent D, while knowledge moves in the opposite direction. That is, patent A is building on patent B, and that is why patent A cites patent B.

We term a first-generation citation as a direct citation of prior work. In our example, these would be A→B, B→C, and C→D. When we discuss the distances over which first-generation citations occur, we are measuring the bilateral distances between these three pairs. We next define a second-generation citation as the culmination of two steps in the citation chain: A→→C and B→→D. When we discuss the distances over which second-generation citations occur, we are measuring the bilateral distances between these end points, removing the intermediate step (A and C, B and D). Our simple example also has a single third-generation citation, A→→→D, and we would measure this distance as the bilateral distance between patents A and D.

Figure 2D continues with figure 2A's example to describe our empirical strategy. We place into this graph the A→B→C→D citation sequence just described. We contrived this example to show a pattern where patent A would never have cited patent D directly according to our model. The distance from A to D is too great for the indicated maximal radius, but the distance can be bridged with the intermediate hops through patents B and C.

In reality, some measure of citations occurs at distances that stretch across the full cluster (just as academics cite others at distances that span the globe). In fact, even if knowledge travels as in our model from patent D to patent A via sites B and C, we might still observe a patent A citing directly patent D (just as academics cite papers directly that they learned about through other papers). So the model's structure cannot be taken so strictly as to say that we should never observe citations at distances of the length between A and D. Nevertheless, we can learn a lot about relative distance of knowledge flows by estimating the relative frequencies of citations by distance. Our model suggests that we should observe a higher frequency of first-generation citations when evaluating the shorter distances within clusters, as direct contact can occur at close proximity. Across longer spans, we should observe both fewer first-generation citations and more later-generation citations, indicative of knowledge transmission through a sequence of overlapping interactions.[14]

FIGURE 3.—LOCAL PATENT TECHNOLOGY HORIZONS



A. Figure plots coefficients from regressions of log patent citation counts between pairwise citing and cited postal codes within 150 miles of each other. Explanatory variables are indicator variables for distance bands with effects measured relative to postal codes 100 to 150 miles apart (unreported bands for 75 to 100 miles resemble 70 to 75 miles). Regressions control for an interaction of log patenting in the pairwise postal codes and log expected citations based on random counterfactuals that have the same technologies and years as true citations. Citations within the same postal code are excluded.

B. See panel A. This figure plots coefficients from regressions of log patent citation counts that employ consolidated distance rings around citing postal codes rather than pairwise combinations of postal codes. Explanatory variables are indicator variables for distance rings with effects measured relative to postal codes 100 to 150 miles (unreported bands for 75 to 100 miles resemble 70 to 75 miles). Regressions control log patenting in the distance ring, log expected citations based on random counterfactuals that have the same technologies and years as true citations, and citing postal code fixed effects. Citations within the same postal code are excluded.

We demonstrate this pattern through some simple estimations illustrated in figures 3A and 3B. For figure 3A, we prepare a data set that contains bilateral pairs of all postal codes that patent during the post-1975 period. To focus on local exchanges, we restrict these postal code pairs to those that are within 150 miles of each other. For each postal code z_1, we then identify the number of citations that it makes to the other paired postal code z_2. To be conservative in our approach, we do not examine interactions within the same postal code and exclude citations that firms make internally among their inventions across postal codes.

With this data set, we empirically model the count of citations that patents in postal code z_1 make of patents in a second postal code z_2 using the general form:

[14] The one exception to this would be if knowledge flows are fully transmissible through the cluster such that any site connected to the cluster receives complete effortless access to the knowledge housed at any site in the cluster. The evidence below suggests that this potential exception is not empirically relevant in this setting.

$$Citations_{z_1 \to z_2} = \exp^{\beta \times d_{z_1,z_2}} (Patents_{z_1} \times Patents_{z_2})^{\gamma},$$

where as before, d_{z_1,z_2} denotes the distance from z_1 to z_2. This expression suggests that citations depend on the interacted stock of patents in the two postal codes and on the distance between the two postal codes, d_{z_1,z_2}. We would anticipate $\beta < 0$ if knowledge flows are declining with distance, and $\gamma > 0$ if a greater number of patents in the two postal codes provides more opportunities for citations. Rearranging this expression gives

$$\begin{aligned} &\ln(Citations_{z_1 \to z_2}) \\ &\quad = \beta \times d_{z_1,z_2} + \gamma \times \ln(Patents_{z_1} \times Patents_{z_2}), \end{aligned}$$

the starting point for our first estimating equation. We make three further modifications. First, beginning with the citations outcome variable, 0 citations may be observed even where patents exist (and this lack of exchange is important information). Our base estimation thus takes $\ln(1+Citations_{z_1 \to z_2})$ to be the citations outcome variable, and we model other variations below to test the sensitivity of the $\ln(X) \mapsto \ln(1+X)$ transformation.

Second, there are multiple ways that one might define distance to potentially allow for nonlinear effects that our model emphasizes. Our first approach is to estimate the role of distance in a nonparametric format using a series of indicator variables $I(\cdot)$ for distance bands between postal codes. We define a vector of distance bands as within 1 mile (but not the same postal code), (1,3] miles, (3,5] miles, (5,10] miles, (10,15] miles, …, (95,100] miles, and (100,150] miles. We denote the set of distance rings as DR and include separate indicator variables for each distance band up to postal codes being (95,100] miles apart. Our β_{dr} coefficients will thus measure the difference in citation rates observed for a distance interval compared to the reference category of being more than 100 miles apart in the technology cluster.

Finally, the $\ln(Patents_{z_1} \times Patents_{z_2})$ control is important, but it is also weak. Our initial tests include all patents, and the patents in the two postal codes may be from very different fields. Thus, while raw citation counts display excessive localization, they may appear localized simply because different types of patenting firms are clustered together. To model this underlying landscape in the most flexible way possible, we generate random citation pairs comparable to our observed sample. For every patent that is actually cited, we randomly draw a counterfactual patent from the pool of all patents with the same technology class and application year as the true citation. This method has been used extensively in the literature, and we make two modifications that reflect our sample design. First, we exclude other patents of the citing firm from the pool of potential draws, just as we exclude within-firm citations in the primary sample. Second, we build the pool of potential patents using only patents within a 250-mile radius of the citing patent. We do not exclude the original cited patent from the random draws, and thus we use the original citation if there are no other patents with the same technology and application year in the defined spatial radius. Relative to simple patent counts, this counterfactual distribution has the advantage of very closely matching the underlying properties of local inventions and their technological foundations; it is a much stronger control, for example, than using simple patent counts.

With these three adjustments, our core estimating equation for figure 3A becomes,

$$\begin{aligned} &\ln(1 + Citations_{z_1 \to z_2}) \\ &\quad = \alpha + \sum_{dr \in DR} \beta_{dr} \times I(d_{z_1,z_2} = dr) \\ &\qquad + \gamma \times \ln(Patents_{z_1} \times Patents_{z_2}) \\ &\qquad + \eta \times \ln(1 + Expected\ Citations_{z_1 \to z_2}) + \varepsilon_{z_1 \to z_2}. \end{aligned} \tag{1}$$

The solid line in figure 3A plots the β_{dr} coefficients for first-generation citations like the A→B example discussed in figure 2D. First-generation citations are quite concentrated at short distances and decline almost monotonically with increasing distance. The citation premium loses half of its strength by (15,20] miles, and postal codes that are 40 miles or more apart are very similar to those in the reference category of being 100 to 150 miles apart. This substantial decay echoes very closely the localized networking results of Arzaghi and Henderson (2008) and the spillover estimations of Rosenthal and Strange (2003, 2008). It is important to recall that we have excluded interactions within the same postal code in order to be conservative. The within–postal code citation premium is larger in magnitude than that observed for neighboring postal codes within 1 mile of each other.

The dashed line in figure 3A graphs the spatial patterns of second and third generations of patent citations, equivalent to the A→→C and A→→→D example. To construct the second-generation citation profile for postal code z_1, we start with the patents that were cited directly by firms in postal code z_1 within a 250-mile radius around postal code z_1. We collect the citations that those patents made to other patents within 250 miles of their postal code. We then calculate the distances from the focal postal code z_1 to these citations, and we will focus again on the second-generation citations that fall within 150 miles of postal code z_1. We take this approach to provide very flexible local distances. Note, for example, that the distance from postal code z_1 to a given second-generation citation can be closer than the first-generation citation that links it. We repeat the same process for third-generation citations. Specification (1) is again used to compare rates in local distances to the rates that exist over 100 to 150 miles.

The results are intuitive and agree with the developed model. At very small distances, later-generation citations are substantially less frequent than first-generation citations. This gap quickly closes, and from distances of 10 to 25 miles, the relative frequencies are very similar. After 25 miles or so, there follows a distance interval where later-generation citations have a greater relative frequency than first-generation citations. These relative differences slowly decay thereafter, and at longer distances, the spatial overlaps become very similar across generations.

Figure 3B plots comparable evidence from a second approach. Rather than use bilateral postal code pairs, we sum the activity of postal codes that falls into the distance rings used above. This approach renders our analysis less sensitive to vagaries of postal code mappings and the issues that one encounters with zero citations; the corresponding disadvantage is that we sacrifice some of the granularity that the bilateral estimations allow. The consolidated empirical framework also allows us to include in the estimations a vector of fixed effects ϕ_{z_1} for citing postal codes. These fixed effects remove persistent differences that exist across postal codes in citation counts such that we are exploiting only variation in how much postal code z_1 cites other postal codes in its technology cluster more or less than typical for postal code z_1. The second estimating equation takes the form

$$\begin{aligned}\ln(1 + Citations^{Ring}_{z_1 \to dr}) &= \sum_{dr \in DR} \beta_{dr} \times I(d_{z_1,z_2} = dr) + \gamma \times \ln(Patents^{Ring}_{dr}) \\ &\quad + \eta \times \ln(1 + Expected\ Citations^{Ring}_{z_1 \to dr}) \\ &\quad + \phi_{z_1} + \varepsilon_{z_1 \to dr}. \end{aligned} \quad (2)$$

These regressions measure the β_{dr} coefficients relative to the activity observed in the excluded distance ring of 100 to 150 miles apart. The solid and dashed lines in figure 3B again plot the first- and later-generation citations, respectively. At very short distances, first-generation citations show greater relative frequency compared to later-generation citations. The differences reverse at moderate distance ranges.

Online appendix tables 1, 2a, and 2b provide complete details on these estimations and descriptive statistics. Online appendix tables 2a and 2b report very similar results to figures 3A and 3B, respectively, when zero-citation cells are excluded, we drop the expected citations controls, and we include own postal citations.

These differences across citation generations suggest that knowledge flows are not fully transmissible through a cluster, but instead follow a pattern indicated by the Silicon Valley example and our model's structure. In figure 2D, the chain of interconnected hops A→B→C→D aids site A's access to knowledge from sites around sites C and D. Moreover, the extra strength for first-generation citations over very short distances offers an approach to identifying maximal radii of interactions—we investigate this next. While it is important to note that other models may be able to generate these patterns, this framework does provide suggestive evidence on how knowledge movements through clusters conform to our model's structure.

IV. Maximal Radii and Spatial Cluster Patterns

Our theory connects the maximal radius of firm interactions with cluster structure. We illustrate these predictions by looking at differences across 36 technologies using the subcategory level of the USPTO system. Hall et al. (2001) describe these technology groups, and examples include Semiconductors, Optics, and Resins. Similar to the analysis conducted in figures 3A and 3B using all patents, we exploit patent citations separately within these individual technology fields so as to measure their radii of interaction. We then examine whether patterns across technologies' cluster shapes and sizes and our measured radii conform to our model's predictions. This section analyzes predictions individually, and the next section models the predictions jointly using continuous density measurement techniques.

A. *U.S.-Based Maximal Radii*

We proxy the maximal radius of interaction for each technology through the citation localization patterns evident among patents within that technology. One technology, for example, may show that most of the citations that exist within local areas occur across firms with a bilateral distance of 10 miles or less. A second technology's local citations could occur more evenly over distances of 0 to 70 miles. In the context of figures 3A and 3B, this second technology would have a much flatter citation premium for short distances. While we cannot put an exact distance on each technology's maximal radius, we can use the differences across technologies in these observable citation patterns to proxy relative differences in their maximal radii.

Our sample preparation for these estimations is similar to that used for the previous graphs. The sample is again restricted to postal codes that are observed to patent in a technology. To be conservative, we again consider citations that are outside the same postal code, excluding self-citations for firms. We also exclude cases where we believe that an inventor has moved and is self-citing his or her prior work. There are several ways that one can attempt to measure these spillover radii from the data, and we consider three formats. These approaches are all simpler than the flexible estimations undertaken in equations (1) and (2) but similar in spirit. These simpler formats are necessary given the substantial reduction in data points when estimating citation patterns on a technology-by-technology basis (especially when extended to the United Kingdom, as noted below). Online appendix table 3 lists the radii measured by technology.

Our first technique considers each technology q in isolation, measuring its citation decay with distance in a log-linear form,

$$\begin{aligned}\ln(1 + Citations_{q,z_1 \to z_2}) &= \beta_q \times \ln(d_{z_1,z_2}) + \gamma_q \times \ln(Patents_{q,z_2}) \\ &\quad + \phi_{q,z_1} + \varepsilon_{q,z_1 \to z_2} \text{ for all } q. \end{aligned} \quad (3)$$

Thus, we estimate a single β_q parameter for how the rate of citations declines with distance. By estimating only one parameter for distance's role, we greatly increase our empirical power for these technology-level estimations. As we are looking at patents and patent citations only within a single technology, we no longer calculate the random citation counterfactual. The patents themselves capture the underlying technology landscape. These estimations are weighted by an interaction of patent counts in the two postal codes. With this technique, Semiconductor and Electrical Devices show the greatest citation localization (most negative β), while Heating and Apparel and Textiles show the weakest role for distance (β in the neighborhood of 0).

Our second technique makes several changes to equation (3) to ensure robustness of technique. We estimate

$$\begin{aligned}\ln(1 + Citations_{q,z_1 \to z_2}) &= \sum_q \beta_{q,0-10} \times I(d_{z_1,z_2} \leq 10) \\ &+ \sum_q \beta_{q,10-30} \times I(10 < d_{z_1,z_2} \leq 30) \\ &+ \gamma \times \ln(Patents_{q,z_1} \times Patents_{q,z_2}) \\ &+ \phi_q + \varepsilon_{q,z_1 \to z_2}. \end{aligned} \tag{4}$$

The core differences between this approach and equation (3) are that we estimate all of the citation declines jointly so that γ is restricted to be the same across technologies; we return to our indicator variable approach for estimating the role of distances in a more flexible manner; and we include a vector of fixed effects for technologies instead of postal codes. We do not have the data to estimate distance rings as finely grained as those considered in the preceding exercises, so we include only indicator variables for bilateral distances of (0,10] miles and (10,30] miles. Thus, the reference group is bilateral postal code pairs of distances between 30 and 150 miles. Our second measure of technology spillover horizons is the observed premium $\beta_{q,0-10}$ over the first 10 miles compared to the reference group. With this technique, Information Storage and Semiconductors show the greatest citation localization (most positive β), while Furniture and Receptacles show the weakest role for distance (β in the neighborhood of 0). This measure has a 0.7 correlation with that calculated through the equation (3) measure.

Finally, our third approach is completely nonparametric and relies on the relative prevalence of first- versus later-generation citations by distance for technologies (using up to six generations). All technologies start with first-generation citations having the highest relative prevalence, and all technologies eventually at some distance have later-generation citations more prevalent. For each technology, we identify the distance at which this crossing point occurs in 2-mile increments. The series can be jumpy, especially for smaller technologies, so we make the specific requirement that one of two conditions be met: (a) the relative frequency of later-generation citations exceeds first-generation citations by 2% or more, or (b) the relative frequency of later-generation citations exceeds first-generation citations for three consecutive distances. Many technologies show crossing points at 10 miles or less, while Receptacles and Pipes and Joints show the longest crossings at more than 20 miles. Overall, this measure is less correlated with the first two metrics at 0.2 to 0.3.

B. U.K.-Based Maximal Radii

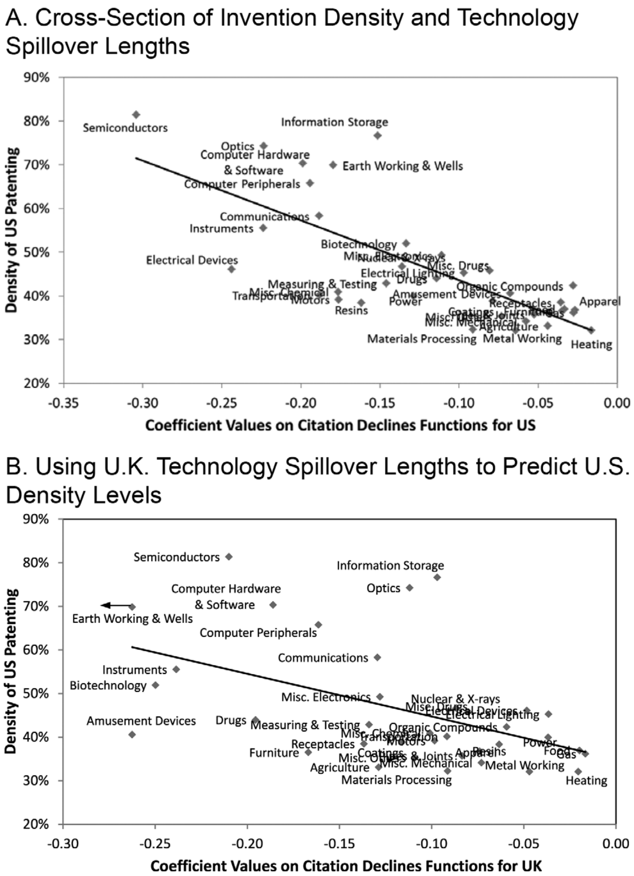
We find evidence of a strong correlation between lengths of microinteractions among firms (within technologies) and their associated cluster shapes and sizes. It is natural to worry in this setting about reverse causality. Existing cluster shapes and economic geography likely influence citation behavior. Moreover, technology clusters may have their spatial locations for unmodeled reasons (e.g., historical accidents, fixed university locations). The length of patent citations could then be determined by the geographical features of these locations.

To address this, we calculate citation premiums similar to our first two metrics using patent data from the United Kingdom. Ellison et al. (2010) introduce this technique and discuss its strengths and limitations. The central idea behind this identification strategy can be illustrated with the semiconductors technology. Many semiconductor firms are located in Silicon Valley, and as the map in figure 1 illustrates, Silicon Valley is circled by water, mountains, and protected land. It could be that the cluster density and short citation ranges that we observed are due to this industry having developed in a location with natural features that pushed it toward density and tight connections. Perhaps if the semiconductors industry had instead grown up in Houston, the industry would not display citation localization. If so, the data would describe features like our model's predictions, but the connection would be spurious.

We can provide a safeguard against these concerns by measuring citation premiums in the United Kingdom, which are not influenced by the local terrain of the United States or similar factors. This test does not solve every potential endogeneity concern, but it certainly provides traction against some of the most worrisome endogeneity. To implement this strategy, we geocode all city names and postal codes associated with U.K. inventors. To provide more accurate city assignments, we also manually search for addresses of firms in the United Kingdom with more than fifty USPTO patents. Calculating bilateral distances among pairwise city combinations, we then estimate a second set of technology-level citation regressions that parallels our U.S. estimations.

The U.K. calculations face several important limitations relative to the U.S. calculations. First, and most important, there are significantly fewer data points to estimate these citation premiums (the U.K. sample is less than one-tenth of the U.S. sample size). Second, the geocoding has greater measurement error, perhaps most concentrated around London, and is coarser than in the United States. As a consequence, we do not attempt to exclude same-region citations as we do for the U.S. data. We also do not attempt to implement our third approach of measuring the crossing point of citation generations, as the data are too sparse with respect to later-generation citations. While these limitations restrict

FIGURE 4.—PATENT CLUSTER DENSITY AND SPILLOVER RADIUS



A. A cross-sectional plot of cluster density and technology spillover lengths. Cluster density is measured through bilateral patent distances in each technology. It is the share of patenting that occurs within 50 miles relative to the share within 150 miles. The horizontal axis measures by technology the log rate of citation decay by distance, controlling for underlying patenting and citing postal codes fixed effects. Longer spillover horizons (i.e., weak decay rates) are associated with less dense clusters. The slope of the trend line is −1.336 (0.226).

B. See figure panel A. Estimates use technology spillover lengths in the United Kingdom to address potential reverse causality where U.S. cluster shapes determine spillover lengths. The outlier, raw decay rate of −0.507 for Earth Working and Wells is capped at the second-highest decay rate. The slope of the trend line is −0.979 (0.293). The slope of the trend line is −0.745 (0.158) without the cap.

our analysis somewhat, the U.K. results in this section and the next provide important confirmation of our model's predictions in a manner that addresses some reverse causality concerns.

Online appendix table 3 lists the U.K. metrics. The correlation between the U.S. and U.K. metrics using our first specification, equation (3), is 0.4. The correlation between the U.S. and U.K. metrics using our second specification, equation (4), is 0.2. The two U.K. metrics have a 0.5 correlation among themselves.

C. *Analyses of Single Predictions*

Figure 4A provides a cross-sectional plot of cluster density and our first proxy for maximal radius by technology that uses log-linear decay rates. Density is measured by the share of bilateral distances among patents for a technology over 0 to 50 miles divided by the share of bilateral distances among patents over 0 to 150 miles. Shares range from 30% to over 80%, with a very high share indicating that patents in the technology are very densely packed in one cluster and then mostly absent until the next cluster. There is a visible association between longer spillover horizons (weaker decay rates that approach 0) and less patent density. On the other hand, technologies that display very rapid decay rates and short technology spillover horizons are tightly clustered. Recall that citation decay rates are calculated controlling for the underlying spatial patent distribution, so this relationship is not mechanical. The slope of the trend line is −1.336 (0.226). Very clearly, some of the industries in information technology show exceptional densities. The slope of the trend line is −0.612 (0.082) when capping the density ratio at 50%.

Figure 4B provides a cross-sectional plot of U.S. cluster density against U.K. citation decay rates. The vertical axis is the same as in figure 4A, but we substitute the U.K. citation decay rates for the horizontal axis's measure of technology spillover ranges. The United Kingdom has an outlier, raw decay rate of −0.507 (Earth Working and Wells); we cap this rate at the second-highest decay rate. There are some material adjustments among some information technology industries in figure 4B compared to figure 4A, with, most noticeably, semiconductors' decay rate not being as steep as we measured in the United States. Nevertheless, a close connection exists between the decay rates for technologies in the United Kingdom and associated cluster density in the United States. The slope of the trend line is −0.979 (0.293); it is even sharper at −0.466 (0.115) when capping the U.S. density ratio at 50%. The slope of the trend line is −0.745 (0.158) without the cap for Earth Working and Wells. These patterns provide confidence that the relationships we identify are not being solely determined by unmodeled factors.

Table 1 continues these analyses of single predictions regarding the size and shape of clusters. Each entry in the table is from a separate regression where the outcome variable is indicated in the column heading, and the five panel headers indicate the metric used to model the maximal radius of interaction. Panels A and D consider the log-linear citation decay rates estimated through technique (3) measured in the United States and United Kingdom, respectively. Panels B and E similarly consider the U.S. and U.K. citation premium observed over 10 miles from estimation (4). Finally, panel C models spillover lengths through the crossing points observed for technologies between first- and later-generation citation frequencies.

To make our estimates easily comparable to each other, we transform variables to have unit standard deviation. We also multiply the raw $\beta_{q,0-10}$ coefficient for panels B and E by −1 so that the predicted signs for table 1's regressions are aligned in the same direction. These regressions exploit variation across the 36 technologies, and we control for the size of the technology using its patent count during the 1975–2009 period. Regressions are unweighted and report robust standard errors. We find very similar patterns when weighting technologies by size.

TABLE 1.—BASIC CLUSTER TRAITS AND MAXIMAL RADIUS ESTIMATIONS

	Size of Clusters			Density of Clusters		
	Mean Distance to Other Patents in MSA from the Dominant Postal Code per Technology Prediction: Longer Distance (1)	Median Distance to Other Patents in MSA from the Dominant Postal Code per Technology Prediction: Longer Distance (2)	Herfindahl Index of Patent Distribution over Postal Codes within MSA for Technology Prediction: Weaker HHI (3)	Share of Patents That Occur within 150 Miles of each Other That Occur within 50 Miles Prediction: Less Dense (4)	Column 4's Measure with the Maximum Density Capped at 50% Prediction: Less Dense (5)	Share of Patents that Occur within 50 Miles of Each Other That Occur within 25 Miles Prediction: Less Dense (6)
	A. Measuring Radius through Log-Linear Patent Citation Decay Functions					
Maximal radius of interaction	0.416	0.392	−0.631	−0.776	−0.758	−0.818
	(0.151)	(0.172)	(0.176)	(0.134)	(0.109)	(0.098)
	B. Measuring Radius through Nonparametric Patent Citation Decay Functions					
Maximal radius of interaction	0.579	0.566	−0.721	−0.851	−0.689	−0.858
	(0.107)	(0.111)	(0.148)	(0.154)	(0.185)	(0.171)
	C. Measuring Radius through Comparing First- versus Later-Generation Citation Distributions					
Maximal radius of interaction	0.241	0.257	−0.151	−0.253	−0.177	−0.238
	(0.093)	(0.104)	(0.132)	(0.113)	(0.112)	(0.099)
	D. Measuring Radius through Log-Linear Patent Citation Decay Functions in United Kingdom					
Maximal radius of interaction	0.427	0.363	−0.361	−0.484	−0.505	−0.390
	(0.103)	(0.121)	(0.179)	(0.160)	(0.137)	(0.162)
	E. Measuring Radius through Nonparametric Patent Citation Decay Functions in United Kingdom					
Maximal radius of interaction	0.163	0.110	−0.229	−0.301	−0.229	−0.254
	(0.148)	(0.158)	(0.181)	(0.193)	(0.173)	(0.187)

Table quantifies the relationship between traits of technology clusters and the maximal radius of interactions for technologies. Dependent variables are indicated by column headings, and panel titles indicate the technique employed to measure the maximal radius. A longer maximal radius is predicted to have larger and less dense clusters. Cluster traits are measured during the 1990–1999 period. The sample includes 36 technologies at the subcategory level of the USPTO classification system. Variables are transformed to have unit standard deviation for interpretation. Estimations are unweighted, control for the number of U.S. patents in the technology, and report robust standard errors.

The first three columns examine the size of clusters, where we have the prediction that a longer maximal spillover radius produces a larger cluster. We take metropolitan statistical areas (MSAs) as the unit of observation, measuring the patenting that occurs within the postal codes of each MSA. In the next section, we consider more flexible techniques that do not depend on MSA definitions, as technology clusters may extend past MSA boundaries or across MSAs. This simple starting point is attractive, however, as it does not depend on the structure of the continuous density techniques.

We identify the leading or dominant postal code per MSA in terms of patent counts by technology. Columns 1 and 2 describe the mean and median distance, respectively, from the dominant postal code to other patents in the MSA by technology. These distances are calculated as the weighted averages of the distances from the dominant postal code using postal code centroids. There is a positive relationship in columns 1 and 2, such that a 1 SD increase in the estimated maximal radius of a technology is associated with a 0.24 to 0.58 SD increase in these mean and median distances when using U.S.-based radii in panels A to C. The estimated elasticity is 0.11 to 0.43 when using U.K.-based radii in panels D and E. Overall, with the exception of weaker performance in panel E, these results highlight that a greater spillover range for a technology is associated with longer mean and median distances within MSAs for the technology's patents.

Column 3 evaluates an alternative metric where we calculate the normalized Herfindahl index of patents over the postal codes in a given MSA by technology. A second way that we might observe a greater size of technology clusters within a given MSA is if the patents for the technology are spread out over more postal codes (a weaker Herfindahl index). This prediction connects with our radii as measured in panels A and B, with very strong elasticities of about −0.63 to −0.73, and with the U.K. log-linear decay function in panel D, with a strong elasticity of −0.36. On the other hand, the support in panels C and E is weak. The coefficient elasticity retains the predicted sign, but the results are not statistically significant.

Columns 4 to 6 shift the focus toward the prediction that clusters with longer spillover radii will be less dense. Column 4 continues with the density metric examined in figures 4A and 4B, where we measure the fraction of bilateral distances between patents that are 150 miles or less apart that are in fact 50 miles or less apart (i.e., count of patents with bilateral distances of 50 miles or less/count of patents with bilateral distances of 150 miles or less). This prediction finds support with all of our metrics. After controlling for the size of technology, the estimated elasticity using U.S.-based radii is 0.25 to 0.85; the U.K.-based elasticities are 0.30 to 0.48. These elasticities are precisely measured. Column 5 shows comparable results when capping density at 50%, and column 6 shows similar patterns when we instead consider the density among patents that are 50 miles or less apart by looking at the fraction of these patents that are 25 miles or less apart.

V. Continuous Density Estimations

Overall, the regressions in table 1 suggest that a longer spillover radius for a technology is associated with larger and less dense clusters. To some degree, of course, the different outcome measures that we model in table 1 are variations on a similar theme. Our six outcomes are also ad hoc in their design, in that we do not have any particular reason to examine, for example, the density over 50 miles compared to the density over 43 or 72 miles. This section provides a joint test of our model's predictions in a more rigorous manner using continuous density estimations. We first introduce the Duranton and Overman (2005) methodology that we use, and then show how the shapes of local technology clusters relate to technology spillover horizons.

A. *Duranton and Overman (2005)*

Our empirical work in large part uses a slight variant of the Duranton and Overman (2005, hereafter DO) metric or its underlying smoothed kernel density. This discussion summarizes the DO methodology to show the connection to our theory. (The empirical appendix in our NBER working paper further describes the DO metric and the empirical modifications required for our specific data sets).

The DO metric considers bilateral distances among establishments in an industry. The central calculation is the spatial density of an industry A through a continuous function:

$$\hat{K}_A(d) = \frac{1}{hN^A(N^A-1)} \sum_{i=1}^{N^A-1} \sum_{i'=i+1}^{N^A} f\left(\frac{d - d_{j(i),j(i')}}{h}\right). \quad (5)$$

Here, as in our basic model setup, $d_{j(i),j(i')}$ is the Euclidean distance between the spatial locations of establishments $j(i)$ and $j(i')$ within industry A. The double summation considers every pairwise bilateral distance within the industry analyzed (i.e., $N^A(N^A-1)/2$ distances). Establishments receive equal weight, and the function f is a gaussian kernel density function with bandwidth h that smooths the series.

The resulting density function provides a distribution of bilateral distances for establishments within an industry. Across all potential distances, ranging from firms being next door to each other to being across the country from each other, this distribution sums to 1. Smoothed density functions are calculated separately for each technology or industry analyzed. Industries where establishments tend to pack together tightly in cities, for example, are measured to have higher densities $\hat{K}_A(d)$ at short distance ranges.

While the density function is of direct interest, it is also important to compare the observed distributions of bilateral distances to general activity in the underlying economy. This comparison provides a basis for saying whether an industry's spatial concentration at a given distance is abnormal. Because the density functions for small industries with fewer plants are naturally lumpier, these comparisons are specific to industry size. Operationally, comparisons are calculated by repeating the density estimation for 1,000 random draws of hypothetical industries of equivalent size to the focal industry A. This procedure, further discussed in the working paper's empirical appendix, provides 5%/95% confidence bands for each industry and distance that we designate as $K_A^{LCI-U}(d)$ and $K_A^{LCI-L}(d)$.

Industry localization γ_A and dispersion ψ_A at distance d are defined using the DO formulas:

$$\gamma_A(d) \equiv \max\left[\hat{K}_A(d) - K_A^{LCI-U}(d), 0\right],$$
$$\psi_A(d) \equiv \begin{cases} \max\left[K_A^{LCI-L}(d) - \hat{K}_A(d), 0\right] & \gamma_A(d) = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

Positive localization is observed when the kernel density exceeds the upper confidence band; similarly, positive dispersion occurs when the kernel density is below the lower confidence band. In between, an industry is said to be neither localized nor dispersed, and both metrics have a 0 value. To allow for consistent and simple graphical presentation, we present a combined measure of localization and dispersion:

$$\gamma_A^C(d) \equiv \gamma_A(d) - \psi_A(d). \quad (7)$$

An industry is neither localized nor dispersed at a given distance if its density is within the 5%/95% confidence bands. In such cases, $\gamma_A^C(d)$ has a value of 0. Excess density at distance d has a positive value, while abnormally low density carries a negative value. Our estimations analyze these local departures in a systematic manner across industries.

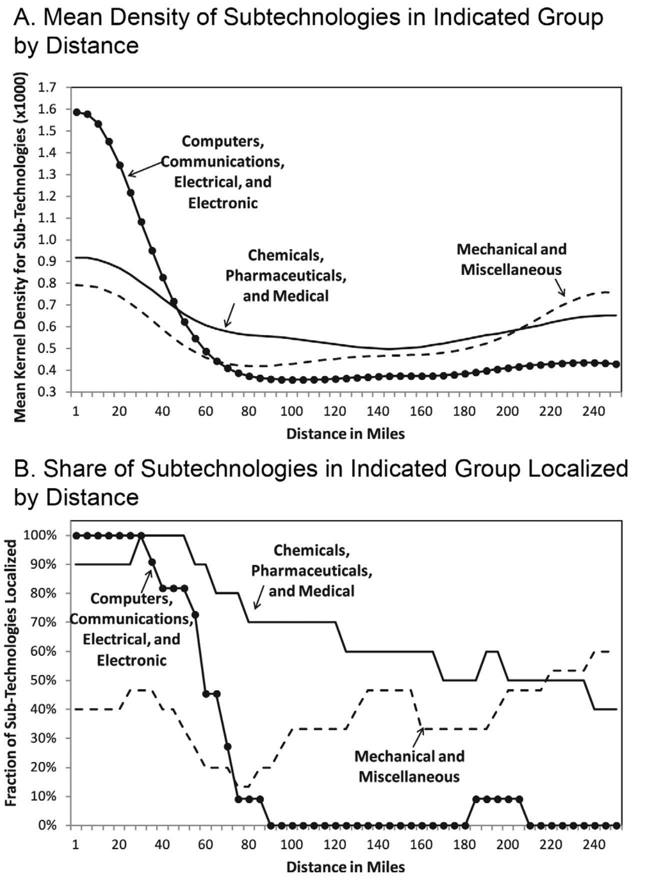
B. *Descriptive Statistics*

For each technology, we estimate the continuous DO spatial density metric described above using patent data from 1990 to 1999.[15] Distances are calculated using postal code centroids. Figures 5A and 5B provide descriptive evidence on patent cluster shapes. We group our 36 technologies into three broad buckets based on the categories of the USPTO system following Hall et al. (2001): Chemicals, Pharmaceuticals, and Medical (categories 1 and 3), Computers, Communications, Electrical, and Electronics (categories 2 and 4), and Mechanical and Miscellaneous (categories 5 and 6).

Figure 5A provides the average kernel density, equation (5), by distance for the technologies that are contained within each grouping. The technologies within the Computers/Electronics grouping show high spatial concentration over the first 30 miles but then exhibit very low density at moderate to long distances. The Chemicals/Medical grouping has lower average density levels at short ranges, but then exhibits the highest average spatial densities over medium

[15] Computational limitations, primarily around constructing the counterfactuals, require that we calculate these densities using patents from 1990 to 1999. We calculate very similar densities for a few smaller technologies when instead considering 1975 to 2009.

FIGURE 5.—PATENT KERNEL DENSITIES AND PATENT LOCALIZATION MEASURES



A. Mean kernel density of technologies by each distance (×1000 for scale). The sample includes 36 subcategories of the USPTO system organized into three simple divisions. Kernel density is calculated using pairwise distances among inventors in a technology.

B. See panel A. Localization is calculated through a comparison of the kernel density estimations for technologies with Monte Carlo confidence bands under the Duranton and Overman (2005) technique. Technologies are considered localized at a distance if they exhibit abnormal density compared to 1,000 random draws of U.S. inventors of a similar size to the technology. Local confidence bands are set at 5%/95% for this determination. Localization looks very similar with 1%/99% confidence bands.

distances. By contrast, the Mechanical/Miscellaneous grouping does not exhibit very strong patterns.

Figure 5B uses the localization metric, equation (6), plotting the fraction of technologies within each group that are localized. Every technology within the Computers/Electronics grouping shows abnormally high spatial concentration over the first 30 miles. After 35 miles, however, localization within this group decays rapidly and is mostly gone by 70 miles. On the other hand, the Chemicals/Medical grouping shows abnormally high spatial concentration over 30 to 60 miles, with a much slower decay rate thereafter. Finally, there is little material variation by distance in the number of technologies localized for the Mechanical/Miscellaneous grouping.

These patterns roughly conform with our predictions, as our measures of technology spillover radii in online appendix table 3 tend to be smaller for Computers/Electronics than for Chemicals/Medical or Mechanical/Miscellaneous. Greater requirements for very close knowledge exchange are visibly associated with shorter, denser spatial clusters across these broad groups. This description, however, does not take advantage of the heterogeneity within groups or the intensity of agglomeration, to which we turn next.[16]

C. *Complete Density Plots*

While transparent, table 1's analyses are incomplete in that they do not describe the full distribution of firm localization behavior. They also do not account for differences in technology size, which can have a mechanical effect on density estimates. We now use the DO methodology to describe these patterns more completely.

We begin with the kernel density $\hat{K}_A(d)$ defined in equation (5) for a technology A. The process of assigning localization (6) involves nonmonotonic transformations of the data, and it is thus useful to view the simpler density functions first. With some abuse of notation, we define $\hat{K}_{A,d}$ as the sum of the kernel density over 5-mile increments starting from 0 to 5 miles and extending to 245 to 250 miles. We again index distance rings with dr and denote the set of distance rings as DR, although the distance rings are different from the citations analysis.

Figures 6A and 6B present coefficients from empirical specifications of the form

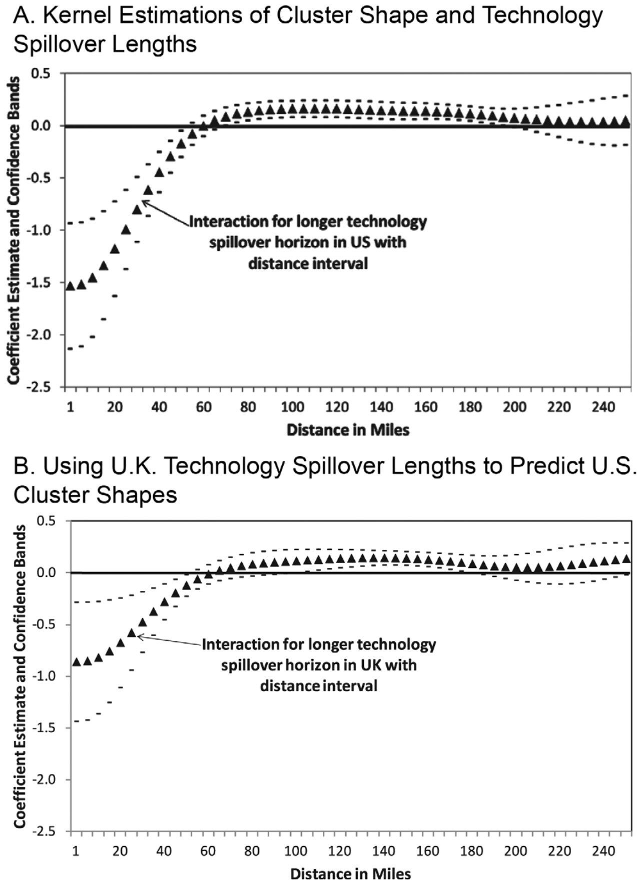
$$\hat{K}_{A,d} = \left(\sum_{dr \in DR} \beta_{dr} \times I(d = dr) \times SpilloverRadius_A \right) + \phi_d + \varepsilon_{A,d}. \quad (8)$$

These estimations provide a continuous description of how technology cluster shapes vary with technology horizons. $SpilloverRadius_A$ is the technology spillover radius for industry A calculated through through our five techniques and listed in online appendix table 3. Greater values of $SpilloverRadius_A$ correspond to longer maximal radii in our model, and we thus anticipate finding larger and less dense clusters for these technologies. We transform $\hat{K}_{A,d}$ and $SpilloverRadius_A$ to have unit standard deviation to aid interpretation, and we evaluate β_{dr} at each distance ring.

A vector of distance fixed effects ϕ_d controls for typical agglomeration densities by distance. They thus directly account for the overall spatial density of patenting so that our estimations consider differences across technologies. As the vector of distances fully contains the support of distances, we do not include a main effect for $SpilloverRadius_A$. Higher values of β_{dr} indicate that technologies with longer spillover radii show greater spatial density at that distance. The cross of 51 distances and 36 technologies yields 1,836 observations per estimation.

[16] At first it may appear odd that a majority of technologies are deemed localized when the confidence bands are selected such that only 5% of the counterfactuals reach them. This is to be expected if agglomerative forces exist, however, as the counterfactuals build on all patent locations. The counterfactuals are not selected such that only 5% of technologies will be deemed agglomerated. This levels effect for localization, along with its overall decline with distance, is predicted by our model if sites are distributed uniformly but agglomerative forces exist for nearly all technologies.

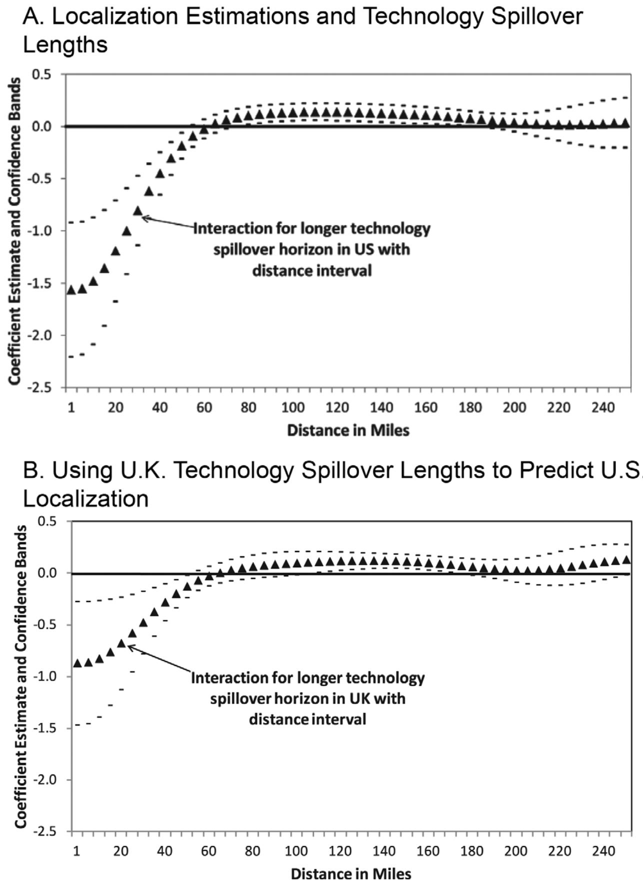
FIGURE 6.—PATENT CLUSTER SHAPE AND SPILLOVER RADIUS



A. Figure plots coefficients from regressions of kernel densities by distance for 36 technologies. Technology decay rates are measured as in figure 4A by the log rate of citation decay by distance, controlling for underlying patenting and citing postal code fixed effects. Regressions include fixed effects for each distance. Dashed lines are 90% confidence bands. Technologies with longer spillover ranges (i.e., weaker decay functions) show lower density at short distances and increased activity over medium distances (i.e., larger and less dense clusters).

B. See panel A. Estimates use technology spillover lengths in the United Kingdom to address potential reverse causality where U.S. cluster shapes determine spillover lengths. Technology decay rates are measured as in figure 4B by the log rate of citation decay by distance in the United Kingdom, controlling for underlying patenting and citing postal code fixed effects.

FIGURE 7.—PATENT LOCALIZATION AND SPILLOVER RADIUS



A. See figure 6A. The dependent variable is updated from the kernel density in figure 6A to be the measurement of localization developed by Duranton and Overman (2005). Technologies are considered localized at a distance if they exhibit abnormal density compared to 1,000 random draws of U.S. inventors of a similar size to the technology. Local confidence bands are set at 5%/95% for this determination. Technologies with longer spillover ranges again exhibit larger and less dense clusters with this technique.

B. See panel A. Estimates use technology spillover lengths in the United Kingdom to address potential reverse causality where U.S. cluster shapes determine spillover lengths.

Figures 6A and 6B present these density estimations using the U.S. and U.K. measures of $SpilloverRadius_A$, respectively, estimated with the log-linear decay rates. Triangles report β_{dr} coefficients. The dashed lines provide 90% confidence bands with standard errors clustered by technology.

Technologies with greater $SpilloverRadius_A$ (i.e., longer maximal radii) are substantially less agglomerated at very short distance horizons. An SD increase in $SpilloverRadius_A$ is associated with a 1.5 SD decrease in the density of activity conducted at 5 miles or closer using the U.S. measure; the U.K.-based estimate is 0.9 SD. By 60 to 75 miles, the abnormal spatial concentration is no longer statistically different from 0.

Technologies with longer $SpilloverRadius_A$ are overrepresented after 75 miles or thereabouts. Using the U.S. estimate of citation density, these clusters show an abnormal density from 80 to 185 miles that is statistically different from 0 at every 5-mile increment. The U.K. estimation shows a similar pattern, although its point estimates are statistically different from 0 for a shorter distance range. In both cases, the point estimates converge to 0 as distances approach 250 miles. At the edge of this spatial scale, differences in maximal radius are not systematically associated with different agglomeration intensities.

These patterns closely match our model and the predictions given in section IIC regarding maximal radii and cluster shapes. Note that the patterns of underrepresentation followed by overrepresentation are not mechanical. Other attributes, for example, could predict higher spatial concentration for a technology at all spatial distances to 250 miles.[17]

Figures 7A and 7B take the next step of calculating localized deviations from technology-specific confidence intervals using equation (7). The patterns are very similar to figures 6A

[17] The kernel density functions (5) sum to 1 over the support of all bilateral distances in the United States, stretching from next door to several thousand miles. This does not materially influence the cluster descriptions we develop here over the first 250 miles.

TABLE 2.—EXTENSIONS ON CONTINUOUS DENSITY ESTIMATIONS

Distance Interval	Measuring Radius Using U.S. Parametric Citation Decays (1)	Column 1 with Weights for Technology Size (2)	Column 1 Excluding Miscellaneous Categories (3)	Column 1 with Bootstrapped Standard Errors (4)	Measuring Radius Using U.S. Nonparametric Citation Decays (5)	Measuring Radius Using First- versus Later-Generation Citations (6)	Measuring Radius Using U.K. Parametric Citation Decays (7)	Measuring Radius Using U.K. Nonparametric Citation Decays (8)
A. Dependent Variable Is Kernel Density by Distance in Unit Standard Deviations								
[0,25]	−1.338 (0.310)	−1.364 (0.353)	−1.354 (0.313)	−1.338 (0.139)	−1.773 (0.206)	−0.430 (0.209)	−0.754 (0.297)	−0.599 (0.370)
(25,50]	−0.469 (0.120)	−0.423 (0.119)	−0.470 (0.123)	−0.469 (0.080)	−0.623 (0.074)	−0.107 (0.095)	−0.286 (0.107)	−0.244 (0.145)
(50,75]	0.032 (0.053)	0.062 (0.062)	0.043 (0.057)	0.032 (0.030)	0.023 (0.050)	0.073 (0.063)	0.016 (0.057)	−0.017 (0.078)
(75,100]	0.146 (0.047)	0.169 (0.057)	0.159 (0.050)	0.146 (0.022)	0.163 (0.060)	0.128 (0.076)	0.103 (0.068)	0.036 (0.091)
(100,125]	0.159 (0.048)	0.184 (0.049)	0.171 (0.050)	0.159 (0.020)	0.193 (0.058)	0.133 (0.073)	0.135 (0.057)	0.044 (0.079)
(125,150]	0.144 (0.048)	0.147 (0.044)	0.156 (0.050)	0.144 (0.021)	0.209 (0.038)	0.116 (0.063)	0.146 (0.042)	0.045 (0.064)
(150, 175]	0.130 (0.043)	0.128 (0.042)	0.144 (0.044)	0.130 (0.019)	0.217 (0.027)	0.097 (0.061)	0.121 (0.043)	0.018 (0.062)
(175,200]	0.091 (0.045)	0.110 (0.047)	0.107 (0.046)	0.091 (0.022)	0.202 (0.031)	0.080 (0.067)	0.070 (0.060)	−0.051 (0.067)
(200,225]	0.050 (0.081)	0.119 (0.071)	0.071 (0.082)	0.050 (0.037)	0.214 (0.059)	0.076 (0.072)	0.064 (0.096)	−0.113 (0.090)
B. Dependent Variable Is Localization Metric Using 5%/95% Confidence Bands by Distance in Unit Standard Deviations								
[0,25]	−1.359 (0.332)	−1.382 (0.376)	−1.375 (0.333)	−1.359 (0.149)	−1.838 (0.213)	−0.444 (0.216)	−0.761 (0.307)	−0.631 (0.387)
(25,50]	−0.476 (0.126)	−0.426 (0.127)	−0.478 (0.128)	−0.476 (0.084)	−0.642 (0.078)	−0.116 (0.098)	−0.289 (0.111)	−0.250 (0.150)
(50,75]	0.010 (0.052)	0.044 (0.057)	0.020 (0.055)	0.010 (0.029)	0.016 (0.045)	0.068 (0.066)	0.000 (0.058)	−0.013 (0.082)
(75,100]	0.121 (0.049)	0.147 (0.055)	0.133 (0.051)	0.121 (0.022)	0.153 (0.055)	0.122 (0.077)	0.085 (0.069)	0.041 (0.094)
(100,125]	0.137 (0.049)	0.164 (0.047)	0.148 (0.050)	0.137 (0.021)	0.184 (0.054)	0.126 (0.072)	0.115 (0.057)	0.044 (0.081)
(125,150]	0.123 (0.050)	0.128 (0.047)	0.133 (0.051)	0.123 (0.022)	0.201 (0.036)	0.108 (0.060)	0.121 (0.043)	0.043 (0.064)
(150, 175]	0.098 (0.046)	0.102 (0.045)	0.109 (0.046)	0.098 (0.019)	0.205 (0.028)	0.088 (0.058)	0.093 (0.042)	0.015 (0.062)
(175,200]	0.054 (0.044)	0.079 (0.044)	0.067 (0.044)	0.054 (0.021)	0.189 (0.036)	0.073 (0.066)	0.045 (0.055)	−0.050 (0.067)
(200,225]	0.019 (0.079)	0.086 (0.071)	0.039 (0.079)	0.019 (0.036)	0.206 (0.063)	0.070 (0.071)	0.043 (0.090)	−0.105 (0.088)

Table quantifies the relationship between the density of technology clusters by distance intervals and the technology's maximal radius of interaction. Panel A considers the kernel density estimates for technologies, and panel B considers the localization metric of Duranton and Overman (2005). The explanatory variables are interactions of indicator variables for distance bands with technology-level spillover lengths. Technologies with longer spillover ranges are predicted to show lower density at short distances and increased activity over medium distances (i.e., larger and less dense clusters). For columns 1 to 4, technology spillover lengths are measured as in figure 4A by the log rate of citation decay by distance, controlling for underlying patenting and citing postal code fixed effects. Variations on technology spillover lengths are employed in columns 5 to 8 similar to table 1. The sample includes 36 technologies at the subcategory level of the USPTO classification system. Variables are transformed to have unit standard deviation for interpretation. Except where noted, estimations are unweighted, control for fixed effects by distance, and report robust standard errors. Online appendix tables 4 to 6 provide additional robustness checks, sample splits, and point-by-point regressions.

and 6B. The lack of density at very short spatial horizons is robustly different from the random counterfactuals and very similar to the kernel plots. The abnormally high spatial concentration at moderate spatial horizons is weaker than in the raw kernel density plots, with the U.S. and U.K. estimators both exhibiting a narrower range where they are statistically different from 0.

Overall, these figures jointly illustrate our central model predictions. A longer maximal radius, or weaker spillover density, is very strongly associated with reduced agglomeration at very short spatial horizons (i.e., the cluster is less dense). These same technologies tend to be overrepresented at moderate spatial horizons (i.e., the clusters are larger). The latter result is very strong in the raw U.S. data, and it is mostly confirmed with the U.K. estimator. Moreover, in all cases the initial decline in bilateral densities from the closest feasible values, predicted by proposition 4, is robustly supported.

D. *Robustness Checks and Extensions*

Table 2 provides robustness checks on our results. Panel A provides estimates using the kernel densities of technologies, and panel B provides estimates using patent localization. To facilitate reporting, we estimate a single parameter per 25-mile distance interval, with spatial densities at 225 to 250 miles serving as the reference group. Column 1 in table 2 repeats figures 6A and 7A under this approach.

Column 2 shows very similar results if weighting technologies by their size, with somewhat greater persistence evident for the abnormal densities observed at moderate distances. Column 3 shows slightly stronger patterns when excluding the five technology groups that are defined as residuals (e.g., Miscellaneous Drugs), where consistent clustering concepts may not apply. Finally, column 4 reports bootstrapped standard errors, showing them to be smaller than the clustered standard errors that we otherwise report.

Columns 5 to 8 show the results with our four maximal radius metrics. While the patterns and levels can be different, we discern three key features from this work. First, all five approaches exhibit the basic joint patterns predicted by the model of a longer technology spillover radius being associated with larger and less dense clusters. Second, the reduced-density prediction is robustly confirmed with results holding and precisely measured over the first 50 miles or thereabouts. Finally, the longer prediction finds more moderate support. It is evident in the patterns of all five measures, but it is not statistically different from 0 across any distance range in column 8. In addition, the exact distance intervals at which the increased density is evident varies somewhat by measure. Thus, we find good confirmation of the longer-cluster prediction, but it is generally just directional in nature.

Similar results are found using three additional specification variants. The first employs the density function (5) and introduces the confidence bands $K_A^{LCI-U}(d)$ and $K_A^{LCI-L}(d)$

as precision controls. The second calculates a global index similar to DO's main metric and then evaluates the gradient of this concentration measure across distances. Finally, the DO confidence bands can be adjusted to a 1%/99% significance level.

Online appendix table 4 provides some sample splits that consider features not emphasized in our model and baseline empirics. We seek to establish the robustness of our results by looking at variations within each subsample to see if similar results hold. The first two columns split the sample by the degree to which patents in the technology cite other patents in the same postal code. We have excluded these own–postal code citations in our maximal radius calculations, and so this sample split uses independent data. Our model's structure does not emphasize the intensity of very local interactions (i.e., the $G(0)$ intercepts) but instead the maximum radii. This sample split tests this feature. The empirical patterns that we emphasize are present in both samples, confirming robustness, with the interesting finding that these patterns are more accentuated in industries with very intense local interactions.

Second, our model and baseline empirics consider technology flows only within the same industry, while the development of new patents often draws from several technology areas.[18] To test the robustness to cross-fertilization of technologies, we split technologies by the share of their patents that go to other technology areas. The relationships that we emphasize in this paper look quite similar in the two halves.

Third, our baseline estimations do not restrict patent citations to be within a specified time interval, but diffusion occurs with time that makes knowledge widely available in a local area and beyond. We anticipate our model's predictions to be more important in industries where access to very recent knowledge is critical. To test this feature, we calculate the share of citations nationally by technology area that occur to patents within the prior five years. The patterns are substantially stronger in the sample of technologies that rely on very recent knowledge, with only the less dense part of the prediction holding in the lower half of the distribution.

Fourth, our model does not include input prices that can generate further sorting across locations by firms. We test how much this feature matters by calculating from the 1990 Census of Populations a weighted average of expected science and engineering wages using the top ten cities for each technology in terms of patent counts. Science and engineering wages are reflective of the wages to be paid to inventors. The patterns are present in both parts of the distribution, with some emphasis toward the technologies developed in areas with above median input costs.[19]

Finally, online appendix tables 5 and 6 provide broader robustness checks on the technique of using continuous density estimation. Rather than undertaking the DO transformations, we simply group observed bilateral distances between patents in technologies that are within 250 miles of each other into a set of distance bins. We then calculate for each technology the fraction of the bilateral distances that fall into each bin. These tables provide the mean and standard deviation of these shares. The first four columns provide breakouts for the first 20 miles at 5-mile intervals, while columns 5 to 12 consider 20-mile increments across the full range to 140 miles.

Online appendix table 5 conducts a set of point-by-point regressions on the shares of patents by technology that fall into each distance bin. Each panel provides a simple set of regressions with each of our techniques for measuring spillover radii. Radius measures are normalized to have unit standard deviation, and we control for the number of patents in the technology. We leave the outcome variables in their raw shares since these shares are easy to interpret. The results show that our conclusions are not being driven by the construction of continuous density metrics.[20] With all five radius measures, we again see evidence that a longer spillover radius is associated with larger and less dense clusters. For example, column 1 finds that a 1 SD increase in the spillover radius lowers the share of patenting within [0,5) miles by 1.7% to 4.1% compared to a base of 4.4%. Similarly, column 5 shows that this same radius increase lowers the [0,20) share by 4% to 12% compared to a base of 16.5%. On the other hand, the later columns show an increase in shares at longer ranges. There are several advantages of employing the DO technique, but these estimates show that our conclusions are robust to variations on this approach.

On a related note, online appendix table 6 reports similar point-by-point regressions where we consider each patent assignee as a single observation. Unassigned patents, which represent about a quarter of all patents, are also retained. Our baseline estimations consider bilateral distances between patents, similar to the employment-weighted estimations of DO. This extension shows quite similar patterns when instead considering bilateral distances among unweighted assignees and individual inventors.

[18] The view stressing industrial concentration is most often associated with Marshall, Arrow, and Romer (MAR). The MAR model emphasizes the benefits of concentrated industrial centers, particularly citing the gains in increasing returns and learning-by-doing that occur within industries. The second view, often associated with Jacobs (1970), argues that major innovations come when the ideas of one industry are brought into a new industrial sector. This perspective stresses that a wealth of industrial diversity is needed to create the cross-fertilization that leads to new ideas and entrepreneurial success. Duranton and Puga (2001) formalize theoretical foundations for this model.

[19] In addition to these four sample splits, we find very similar results to our baseline estimations when we include four single control variables for these dimensions.

[20] There are several key differences of the point-by-point regressions compared to the DO estimations. These raw shares are not smoothed, and they are not being measured relative to confidence intervals. The shares are also constrained to sum to 100% over the 250-mile range, which is not imposed on the continuous density estimates.

VI. Conclusion

This paper introduces a new model of location choice and agglomeration behavior. From a simple and general framework, we show that agglomeration clusters generally cover a substantially larger area than the microinteractions on which they build. In turn, agglomerative forces with longer microinteractions are associated with fewer, larger, and less dense clusters. The theory thereby provides a basis for the use of continuous agglomeration metrics that build on bilateral distances among firms. The theory also rationalizes the use of observable cluster shapes and sizes to rank-order the lengths of underlying agglomerative forces. We find confirmation of our theoretical predictions using variation across patent technology clusters.

We hope that our theoretical framework proves an attractive model for incorporating additional factors that influence firm location and agglomeration behavior. Important extensions include modeling the dynamics of industry life cycles, incorporating interactions across firms in different industries, and incorporating the development of new sites. We believe our setting is an attractive laboratory for structural modeling that would enable recovery of the underlying lengths of microinteractions. These parameters could in turn be useful for understanding spillover transmissions in networks and studying spatial propagation of economic shocks.

We have applied our framework to describing patent technology clusters, but we believe that many more applications in industrial agglomeration are possible. For example, future work could look to price the marginal sites of clusters or identify spillover lengths by examining the location decisions of marginal entrants. Our framework highlights the important information that is contained in those agents' indifference conditions if properly identified. As important, we believe, our framework describes interactions in many other contexts as well. For example, studies find that knowledge flows within firms or universities are substantially shaped by the physical layout of facilities (Liu, 2010). We hope that future work similarly analyzes parallel situations where costs of interaction generate maximal radii.

REFERENCES

Aarland, Kristin, James Davis, J. Vernon Henderson, and Yukako Ono, "Spatial Organization of Firms: The Decision to Split Production and Administration," *RAND Journal of Economics* 38:2 (2007), 480–494.

Akcigit, Ufuk, and William Kerr, "Growth through Heterogeneous Innovations," NBER working paper 16443 (2010).

Alcacer, Juan, and Wilbur Chung, "Location Strategies and Knowledge Spillovers," *Management Science* 53:5 (2007), 760–776.

Alfaro, Laura, and Maggie Chen, "The Global Agglomeration of Multinational Firms," HBS working paper 10-043 (2010).

Arzaghi, Mohammad, and J. Vernon Henderson, "Networking off Madison Avenue," *Review of Economic Studies* 75:4 (2008), 1011–1038.

Audretsch, David, and Maryann Feldman, "R&D Spillovers and the Geography of Innovation and Production," *American Economic Review* 86:3 (1996), 630–640.

Balasubramanian, Natarajan, and Jagadeesh Sivadasan, "What Happens When Firms Patent? New Evidence from US Economic Census Data," this REVIEW 93:1 (2011), 126–146.

Barlet, Muriel, Anthony Briant, and Lauire Crusson, "Location Patterns of Services in France: A Distance-Based Approach," *Regional Science and Urban Economics* 43:2 (2013), 338–351.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?" *Quarterly Journal of Economics* 122:2 (2007), 775–805.

——— "Changes in Transportation Infrastructure and Commuting Patterns in U.S. Metropolitan Areas, 1960–2000," *American Economic Review Papers and Proceedings* 100:2 (2010), 378–382.

Behrens, Kristian, Gilles Duranton, and Frederic Robert-Nicoud, "Productive Cities: Sorting, Selection, and Agglomeration," *Journal of Political Economy* 122 (2014), 507–533.

Billings, Stephen, and Erik Johnson, "A Nonparametric Test for Industrial Specialization," *Journal of Urban Economics* 71:3 (2012), 312–331.

Bleakley, Hoyt, and Jeffrey Lin, "Portage: Path Dependence and Increasing Returns in U.S. History," *Quarterly Journal of Economics* 127 (2012), 587–644.

Breschi, Stefano, and Francesco Lissoni, "Mobility of Skilled Workers and Co-Invention Networks: An Anatomy of Localized Knowledge Flows," *Journal of Economic Geography* 9:4 (2009), 439–468.

Carlino, Gerald, Jake Carr, Robert Hunt, and Tony Smith, "The Agglomeration of R&D Labs," Federal Reserve Bank of Philadelphia working paper 12-22 (2012).

Carlino, Gerald, Satyajit Chatterjee, and Robert Hunt, "Urban Density and the Rate of Invention," *Journal of Urban Economics* 61:3 (2007), 389–419.

Ciccone, Antonio, and Robert Hall, "Productivity and the Density of Economic Activity," *American Economic Review* 86:1 (1996), 54–70.

Dauth, Wolfgang, "The Mysteries of the Trade: Employment Effects of Urban Interindustry Spillovers," IAB discussion paper 15/2010 (2010).

Delgado, Mercedes, Michael Porter, and Scott Stern, "Convergence, Clusters and Economic Performance," *Research Policy* 43 (2014), 1785–1789.

Dempwolf, C. Scott, "Network Models of Regional Innovation Clusters and Their Influence on Economic Growth," Ph.D. dissertation, University of Maryland (2012).

Desmet, Klaus, and Esteban Rossi-Hansberg, "Spatial Development," *American Economic Review* 104 (2014), 1211–1243.

Duranton, Gilles, "Urban Evolutions: The Fast, the Slow, and the Still," *American Economic Review* 97:1 (2007), 197–221.

Duranton, Gilles, and Henry Overman, "Testing for Localization Using Micro-Geographic Data," *Review of Economic Studies* 72 (2005), 1077–1106.

——— "Exploring the Detailed Location Patterns of UK Manufacturing Industries Using Microgeographic Data," *Journal of Regional Science* 48:1 (2008), 313–343.

Duranton, Gilles, and Diego Puga, "Nursery Cities: Urban Diversity, Process Innovation, and the Life Cycle of Products," *American Economic Review* 91 (2001), 1454–1477.

——— "Micro-Foundations of Urban Agglomeration Economies" (pp. 2063–2117), in J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, vol. 4 (Amsterdam: North-Holland, 2004).

Ellison, Glenn, and Edward Glaeser, "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach," *Journal of Political Economy* 105 (1997), 889–927.

——— "The Geographic Concentration of Industry: Does Natural Advantage Explain Agglomeration?" *American Economic Review Papers and Proceedings* 89:2 (1999), 311–316.

Ellison, Glenn, Edward Glaeser, and William Kerr, "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns," *American Economic Review* 100 (2010), 1195–1213.

Fallick, Bruce, Charles Fleischman, and James Rebitzer, "Job-Hopping in Silicon Valley: Some Evidence Concerning the Microfoundations of a High-Technology Cluster," this REVIEW 88 (2006), 472–481.

Fu, Shihe, and Stephen Ross, "Wage Premia in Employment Clusters: How Important Is Worker Homogeneity?" *Journal of Labor Economics* 31 (2013), 271–304.

Fujita, Masahisa, and Jacques-François Thisse, "Economics of Agglomeration," *Journal of the Japanese and International Economies* 10:4 (1996), 339–378.

Glaeser, Edward, *Cities, Agglomeration and Spatial Equilibrium* (Oxford: Oxford University Press, 2008).

Glaeser, Edward, and Matthew Kahn, "Decentralized Employment and the Transformation of the American City," NBER working paper 8117 (2001).

Glaeser, Edward, and William Kerr, "Local Industrial Conditions and Entrepreneurship: How Much of the Spatial Distribution Can We Explain?" *Journal of Economics and Management Strategy* 18 (2009), 623–663.

Glaeser, Edward, William Kerr, and Giacomo Ponzetto, "Clusters of Entrepreneurship," *Journal of Urban Economics* 67:1 (2010), 150–168.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti, "Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings," *Journal of Political Economy* 118:3 (2010), 536–598.

Griliches, Zvi, "Patent Statistics as Economic Indicators: A Survey," *Journal of Economic Literature* 28 (1990), 1661–1707.

Hall, Bronwyn, Adam Jaffe, and Manuel Trajtenberg, "The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools," NBER working paper 8498 (2001).

Hanson, Gordon, "Market Potential, Increasing Returns and Geographic Concentration," *Journal of International Economics* 67:1 (2005), 1–24.

Head, Keith, and Thierry Mayer, "The Empirics of Agglomeration and Trade," in J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, vol. 4 (Amsterdam: North-Holland, 2004).

Helsley, Robert, and William Strange, "Matching and Agglomeration Economics in a System of Cities," *Regional Science and Urban Economics* 20 (1990), 189–212.

——— "Coagglomeration, Clusters, and the Scale and Composition of Cities," *Journal of Political Economy* 122 (2014), 1064–1093.

Holmes, Thomas, and Sanghoon Lee, "Economies of Density versus Natural Advantage: Crop Choice on the Back Forty," this REVIEW 94 (2012), 1–19.

Holmes, Thomas, and John Stevens, "Geographic Concentration and Establishment Scale," this REVIEW 84 (2002), 682–690.

Jackson, Matthew, *Social and Economic Networks* (Princeton: Princeton University Press, 2008).

Jacobs, Jane, *The Economy of Cities* (New York: Vintage Books, 1970).

Jaffe, Adam, Manuel Trajtenberg, and Michael Fogarty, "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors," *American Economic Review Papers and Proceedings* 90:2 (2000), 215–218.

Jaffe, Adam, Manuel Trajtenberg, and Rebecca Henderson, "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics* 108:3 (1993), 577–598.

Kerr, William, "Ethnic Scientific Communities and International Technology Diffusion," this REVIEW 90 (2008), 518–537.

——— "Breakthrough Inventions and Migrating Clusters of Innovation," *Journal of Urban Economics* 67:1 (2010) 46–60.

Kerr, William, and Shihe Fu, "The Survey of Industrial R&D—Patent Database Link Project," *Journal of Technology Transfer* 33 (2008), 173–186.

Krugman, Paul, *Geography and Trade* (Cambridge, MA: MIT Press, 1991).

Liu, Christopher, "A Spatial Ecology of Structure Holes: Scientists and Communication at a Biotechnology Firm," *Management Proceedings* (August 2010), 1–6.

Lucas, Robert, and Esteban Rossi-Hansberg, "On the Internal Structure of Cities," *Econometrica* 70 (2002), 1445–1476.

Lychagin, Sergey, Joris Pinkse, Margaret Slade, and John Van Reenen, "Spillovers in Space: Does Geography Matter?" NBER working paper 16188 (2010).

Marcon, Eric, and Florence Puech, "Evaluating the Geographic Concentration of Industries Using Distance-Based Methods," *Journal of Economic Geography* 3:4 (2003), 409–428.

Marshall, Alfred, *Principles of Economics* (London: Macmillan, 1920).

Marx, Matt, and Jasjit Singh, "More Than Just Distance? Borders and Institutions in Knowledge Spillovers," working paper (2012).

Maurel, Françoise, and Béatrice Sédillot, "A Measure of the Geographic Concentration in French Manufacturing Industries," *Regional Science and Urban Economics* 29 (1999), 575–604.

Menon, Carlo, "The Bright Side of Gerrymandering: An Enquiry on the Determinants of Industrial Agglomeration in the United States," working paper (2009).

Mori, Tomoya, Koji Nishikimi, and Tony Smith, "A Divergence Statistic for Industrial Localization," this REVIEW 87 (2005), 635–651.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura, "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach," this REVIEW 96 (2015), 967–985.

Olson, Gary, and Judith Olson, "Mitigating the Effects of Distance on Collaborative Intellectual Work," *Economics of Innovation and New Technology* 12:1 (2003), 27–42.

Partridge, Mark, Dan Rickman, Kamar Ali, and Rose Olfert, "Agglomeration Spillovers and Wage and Housing Cost Gradients across the Urban Hierarchy," *Journal of International Economics* 78:1 (2009), 126–140.

Pe'er, Aviad, and Ilan Vertinsky, "Survival-Enhancing Strategies of De Novo Entrants in Clusters and Dispersal," working paper (2009).

Porter, Michael, *The Competitive Advantage of Nations* (New York: Free Press, 1990).

Romer, Paul, "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94 (1986), 1002–1037.

Rosenthal, Stuart, and William Strange, "The Determinants of Agglomeration," *Journal of Urban Economics* 50 (2001), 191–229.

——— "Geography, Industrial Organization, and Agglomeration," this REVIEW 85 (2003), 377–393.

——— "Evidence on the Nature and Sources of Agglomeration Economies" (pp. 2119–2171), in J. Vernon Henderson and Jacques-François Thisse, eds., *Handbook of Regional and Urban Economics*, vol. 4 (Amsterdam: North-Holland, 2004).

——— "The Attenuation of Human Capital Spillovers," *Journal of Urban Economics* 64 (2008), 373–389.

Rozenfeld, Hernán, Diego Rybski, Xavier Gabaix, and Hernán Makse, "The Area and Population of Cities: New Insights from a Different Perspective on Cities," *American Economic Review* 101 (2011), 2205–2225.

Saiz, Albert, "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics* 125 (2010), 1253–1296.

Sarvimäki, Matti, "Agglomeration in the Periphery," SERC discussion paper 0080 (2010).

Saxenian, AnnaLee, *Regional Advantage: Culture and Competition in Silicon Valley and Route 128* (Cambridge, MA: Harvard University Press, 1994).

Thompson, Peter, "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-Added Citations," this REVIEW 88 (2006), 383–388.

Thompson, Peter, and Melanie Fox-Kean, "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review* 95 (2005), 450–460.

Zucker, Lynne, Michael Darby, and Marilynn Brewer, "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises," *American Economic Review* 88:1 (1998), 290–306.